VIDEO CODING WITH MC-EZBC AND REDUNDANT-WAVELET MULTIHYPOTHESIS

Joseph B. Boettcher and James E. Fowler
Department of Electrical and Computer Engineering
GeoResources Institute, Mississippi State ERC
Mississippi State University, Mississippi State, MS

ABSTRACT

Motion compensation with redundant-wavelet multihypothesis, in which multiple predictions that are diverse in transform phase contribute to a single motion estimate, is deployed into the fully scalable MC-EZBC video coder. The bidirectional motion-compensated temporal-filtering process of MC-EZBC is adapted to the redundant-wavelet domain, wherein transform redundancy is exploited to generate a phase-diverse multihypothesis prediction of the true temporal filtering. Noise not captured by the motion model is substantially reduced, leading to greater coding efficiency. In experimental results, the proposed system exhibits substantial gains in rate-distortion performance over the original MC-EZBC coder for sequences with fast or complex motion.

1. INTRODUCTION

With the advent of motion-compensated temporal filtering (MCTF), video-coding systems have been able to achieve a high degree of spatial and temporal scalability. MCTF avoids the motion-compensation feedback loop present in traditional video-coding architectures by processing consecutive video frames with temporal transforms in the direction of predicted motion. Accurate motion estimation is important to the success of MCTF, since filtering across poorly matched regions can result in low-quality temporal subbands with "ghosting" artifacts [1].

In [2], Chen and Woods introduced bidirectional MC-EZBC, a fully scalable video coder employing block-based, spatial-domain MCTF to yield state-of-the-art performance. In the MC-EZBC system, motion estimation is carried out via hierarchical, variable-size block matching (HVSBM) in which motion vectors are determined for large blocks of pixels in low-resolution frames then refined for smaller subblocks at successively higher levels of detail. After the motion field is determined, pixels without a one-to-one motion-field connection (the so-called "unconnected" pixels) are located and processed separately. To avoid inefficient temporal filtering, MCTF is performed only if fewer than half the pixels between a pair of consecutive frames are unconnected. If this criterion is met, then the frames are temporally filtered, resulting in highpass and lowpass temporal subbands. This process takes place for each pair of frames in a group of pictures (GOP), after which it is performed recursively on the lowpass temporal subbands. After MCTF is complete, the resulting temporal subbands go through spatial wavelet analysis, completing the 3D decomposition. Finally, the 3D subbands are encoded with the EZBC coder [3].

Since uncertainty is inherent in motion estimation, many video-coding systems use a combination of motion predictions, a concept known as multihypothesis motion compensation (MHMC) [4]. The MC-EZBC system already employs two forms of MHMC—fractional-pixel accuracy, a form of spatial-diversity MHMC made possible by the lifting implementation of MCTF, and bidirectional MCTF, a form of temporal-diversity MHMC in which the multiple predictions come from different frames. However, it is possible to enhance the MC-EZBC system with the addition of a third form of MHMC—multihypothesis prediction via transform-phase diversity. Specifically, we propose a system in which redundant-wavelet multihypothesis (RWMH) [5, 6] is embedded within the MC-EZBC framework. Taking place in the redundant-wavelet domain, MCTF in the proposed system benefits from multiple motion predictions that are diverse in transform phase. This results in more accurate motion compensation, leading to higher-quality temporal subbands and more efficient coding.

We expound upon our approach in the following sections. First, an overview of RWMH is provided in Sec. 2. In Sec. 3, our RWMH-EZBC system is described in detail. Experimental results are presented in Sec. 4, followed by concluding remarks in Sec. 5.

2. REDUNDANT-WAVELET MULTIHYPOTHESIS (RWMH)

The redundant discrete wavelet transform (RDWT) [7, 8] is an approximation to the continuous wavelet transform that removes the downsampling operation from the traditional critically sampled discrete wavelet transform (DWT) to produce an overcomplete representation. The well-known shift variance of the DWT arises from its use of downsampling, while the RDWT is shift invariant since the spatial sampling rate is fixed across scale. Numerous RDWT-based video-

This work was funded in part by the National Science Foundation under Grant No. CCR-0310864.

In *Proceedings of the International Conference on Image Processing*, vol. 3, Genoa, Italy, September 2005, pp. 229–232.

coding systems have been developed, originating with the work of Park and Kim [9]. In most of these systems, the redundancy inherent in the RDWT is used exclusively to permit motion estimation and compensation in the wavelet domain by overcoming the shift variance of the critically sampled DWT. In [5], an entirely new use for the redundancy in the RDWT was presented; specifically, transform redundancy was employed to yield multiple predictions of motion that were combined into a single multihypothesis prediction. This approach represented a new paradigm in MHMC wherein diversity in transform phase yields multihypothesis predictions that enhance motion-compensation performance.

A J-scale RDWT can be considered to be composed of 4^J distinct critically sampled transforms, each corresponding to the choice between even- and odd-phase subsampling in both the horizontal and vertical directions at each scale of decomposition. In the RWMH paradigm, wherein motion estimation and compensation take place in the redundant-wavelet domain, each one of these critically sampled transforms "views" motion from a different perspective and thus forms an independent hypothesis of the true motion of the video sequence. After motion compensation is complete, a multiple-phase inverse RDWT combines these multiple hypotheses into a single prediction.

In [5], a video-coding system is described that incorporates RWMH into the motion-compensation feedback loop of the traditional hybrid, block-based video-coding architecture. An in-depth analysis [6] of this hybrid RWMH architecture reveals that the performance gains over single-phase prediction are largely based on the ability of RWMH to reduce the variance of the prediction residual. That is, noise in the RDWT domain undergoes a substantial reduction in variance when the multiple-phase inverse RDWT is applied, which is due to the well-known fact that the inverse RDWT is a pseudo-inverse operation and thereby consists of a projection onto the range space of the forward transform. Consequently, noise not captured by the motion model is greatly reduced in the hybrid RWMH system, leading to substantial reduction in the variance of the prediction residual in the motion-compensation feedback loop and higher coding efficiency. Additionally, in [10, 11], the RWMH concept was introduced into a general, mesh-based MCTF framework to produce a fully scalable video coder (3D-RWMH). Here, we deploy RWMH into the block-based MCTF framework of MC-EZBC, producing the proposed RWMH-EZBC system.

3. THE RWMH-EZBC SYSTEM

In the MC-EZBC system, motion estimation and temporal filtering take place with the video frames in the original spatial domain, resulting in a single-phase prediction of motion. In the RWMH-EZBC system, we instead perform MCTF in the redundant-wavelet domain in order to generate multiple predictions of motion that are diverse in transform phase. The block diagram for the encoder of our RWMH-EZBC system is shown in Fig. 1.

In Fig. 1, each frame of the input GOP is decomposed with a spatial RDWT, and the resulting frames of RDWT coefficients are used in a bidirectional block-matching motion-vector search. In a J-scale RDWT decomposition, each $B \times B$ block in the original spatial domain corresponds to $3J + 1$ blocks of the same size, one in each subband. We call the collection of these co-located blocks a set; each set contains all the different phases of RDWT coefficients. In the motion-estimation procedure, block matching is used to determine the motion of each set as a whole. As in MC-EZBC, we use HVSBM for motion estimation, adding a cross-subband distortion measure as the matching criterion. Absolute errors for each block of the set are summed such that the coefficients from all phases in both the current and reference frames contribute to the distortion measurement. Specifically, the motion vector for the set located at $[x, y]$ is

$$(d_x, d_y) = \arg \min_{-W \leqslant d_x, d_y \leqslant W} \mathrm{MAE}(x, y, d_x, d_y), \quad (1)$$

where the mean absolute error (MAE) is

$$\mathrm{MAE}(x, y, d_x, d_y) = \frac{1}{B^2} \sum_{k=0}^{B-1} \sum_{l=0}^{B-1} \mathrm{AE}(x + k, y + l, d_x, d_y), \quad (2)$$

and the absolute error (AE) is

$$\begin{aligned} \mathrm{AE}(x, y, d_x, d_y) = {} & 2^{-J} \Big| B_J[x, y, t] - B_J[x + d_x, y + d_y, t - 1] \Big| + \\ \sum_{j=1}^{J} 2^{-j} \Bigg(& \Big| V_j[x, y, t] - V_j[x + d_x, y + d_y, t - 1] \Big| + \\ & \Big| H_j[x, y, t] - H_j[x + d_x, y + d_y, t - 1] \Big| + \\ & \Big| D_j[x, y, t] - D_j[x + d_x, y + d_y, t - 1] \Big| \Bigg). \end{aligned} \quad (3)$$

In the above equations, B_j, H_j, V_j, and D_j are the baseband, horizontal, vertical, and diagonal RDWT subbands, respectively, at scale j. A window $[-W, W]$ is used for the block search.

After the motion field is generated, the number of unconnected pixels between two consecutive frames is calculated. If fewer than half the pixels are unconnected, then temporal filtering takes place between the RDWT frames, with each subband using the same motion field for motion compensation. The same process is carried out for each pair of frames in the GOP, after which it is performed recursively on the lowpass temporal subbands. Once the temporal decomposition of the GOP is complete, an inverse spatial

RDWT is performed on each temporal subband, transforming the coefficients back into the spatial domain. Since each RDWT phase forms an independent hypothesis about the temporal filtering based on its unique perspective, the inverse RDWT implicitly combines these hypotheses into a multihypothesis estimate of what the true temporal filtering should be. At this point, the MC-EZBC system continues as usual, with a 2D spatial wavelet decomposition of the temporal subbands followed by EZBC encoding of the resulting spatio-temporal coefficients.

4. RESULTS

In our experiments, we code the grayscale sequences shown in Table 1 with both the MC-EZBC and RWMH-EZBC systems. Both systems use Haar filters for bidirectional MCTF, while RWMH-EZBC uses the popular 9-7 biorthogonal filter with symmetric extension to perform the spatial RDWT. A GOP size of 8 frames was used, allowing up to 3 levels of temporal decomposition if MCTF is performed at every level. Additionally, the RWMH-EZBC system uses a 1-level spatial RDWT decomposition. All sequences were coded with quarter-pixel motion accuracy.

Average PSNR results for all the test sequences at a fixed rate are provided in Table I. In each sequence, the multihypothesis MCTF employed by the RWMH-EZBC system provided performance gains over MC-EZBC, albeit in varying degrees. The greatest gains were witnessed for sequences with fast or complex motion, such as the "Football" sequence, for which the average PSNR improved on the order of 0.5 dB. For sequences with little motion, such as the "Susie" sequence, the performance gains were not as substantial. The rate-distortion curves in Figs. 2–4 indicate that these observations hold over a range of rates.

5. CONCLUSIONS

In this paper, we deploy RWMH into the prominent fully scalable MC-EZBC video coder by performing MCTF in the domain of the redundant, or overcomplete, wavelet transform. In doing so, we take advantage of the redundancy of the transform to provide multiple predictions of motion that are diverse in transform phase, with each phase "viewing" motion from a different perspective. After MCTF is performed in the redundant wavelet domain, an inverse RDWT transforms coefficients back into the spatial domain, implicitly combining the multiple temporal filterings into a single, multihypothesis prediction of the true temporal filtering.

Experimental results show that the proposed RWMH-EZBC system improves upon the original MC-EZBC. Although we witnessed improved average PSNR for all test sequences, we note that the performance gains associated with RWMH-EZBC are more substantial when the video sequence contains fast or complex motion. This is as expected, as RWMH was more effective for these sequences in the systems of [5, 6, 10, 11] as well. The analysis of [6, 11] reveals that the noise reduction provided by the multiple-phase inverse RDWT of the RWMH process is more effective for these sequences since more noise is left uncaptured by the motion model when motion is fast or complex.

6. REFERENCES

[1] A. Secker and D. Taubman, "Highly scalable video compression using a lifting-based 3D wavelet transform with deformable mesh motion compensation," in *Proceedings of the International Conference on Image Processing*, vol. 3, Rochester, NY, September 2002, pp. 749–752.

[2] P. Chen and J. W. Woods, "Bidirectional MC-EZBC with lifting implementation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 10, pp. 1183–1194, October 2004.

[3] S.-T. Hsiang and J. W. Woods, "Embedded image coding using zeroblocks of subband/wavelet coefficients and context modeling," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 3, Geneva, Switzerland, May 2000, pp. 662–665.

[4] 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*, vol. 5, Minneapolis, MN, April 1993, pp. 437–440.

[5] 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*, vol. 2, Barcelona, Spain, September 2003, pp. 53–56.

[6] ——, "Motion compensation via redundant-wavelet multihypothesis," *IEEE Transactions on Image Processing*, submitted March 2004, revised February 2005.

[7] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," in *Wavelets: Time-Frequency Methods and Phase Space*, J.-M. Combes, A. Grossman, and P. Tchamichian, Eds. Berlin, Germany: Springer-Verlag, 1989, pp. 286–297.

[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. Berlin, Germany: Springer-Verlag, 1989, pp. 298–304.

[9] 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.

[10] 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*, vol. 2, Barcelona, Spain, September 2003, pp. 755–758.

[11] ——, "3D video coding with redundant-wavelet multihypothesis," *IEEE Transactions on Circuits and Systems for Video Technology*, submitted July 2003, revised April 2004 and March 2005.

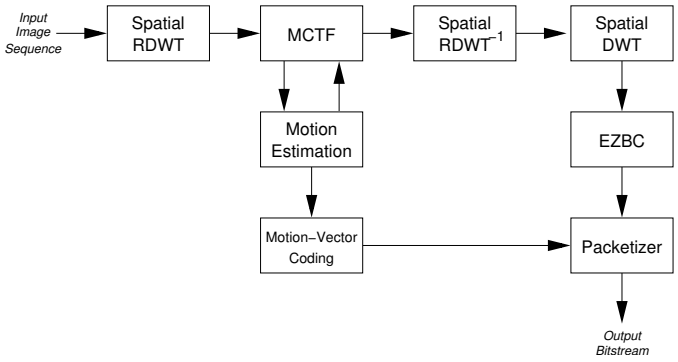

Figure 1: Block diagram of the RWMH-EZBC video-coding system.

Table 1: Distortion averaged over all frames of the sequence for rate of 0.5 bpp.

	PSNR (dB)	
	MC-EZBC	RWMH-EZBC
Football†	29.7	30.2
Table Tennis	36.1	36.4
Foreman	39.6	40.0
Susie†	42.9	43.0
Coastguard	33.5	33.6
NYC†	40.4	40.6

Sequences are CIF (352×288) at 30 Hz except †, SIF (352×240).

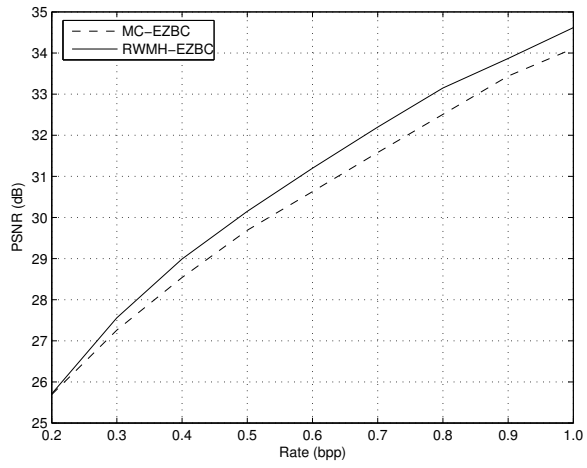

Figure 2: Rate-distortion performance for "Football" at 1/4 pixel accuracy

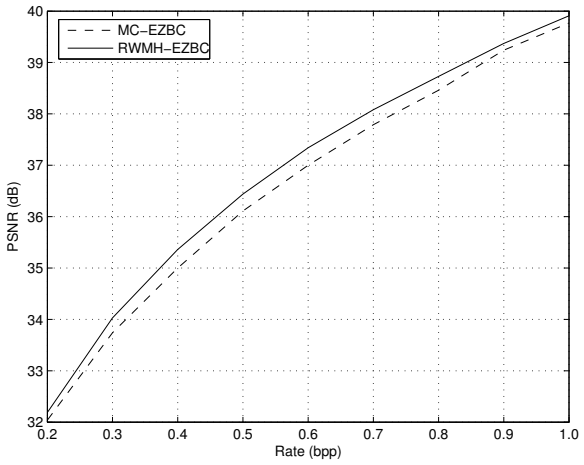

Figure 3: Rate-distortion performance for "Table Tennis" at 1/4 pixel accuracy

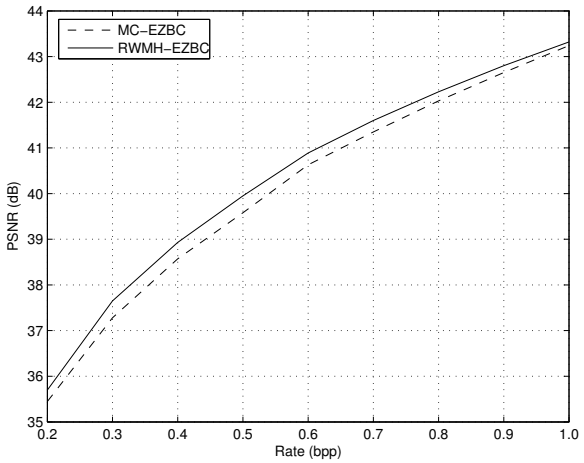

Figure 4: Rate-distortion performance for "Foreman" at 1/4 pixel accuracy