TEMPORALLY SCALABLE MOTION COMPENSATED ADAPTIVE TEMPORAL SUBBAND CODING OF VIDEO

L.L. Winger and A.N. Venetsanopoulos

Department of Electrical and Computer Engineering
University of Toronto
Toronto, Ontario, M5S 3G4, Canada

ABSTRACT

It has been asserted that temporal subband coding (TSB) is inferior to predictive coding for regionally motion compensated (e.g. block-based MC) temporally scalable compressed video [1]. There are two major disadvantages of TSB coding: temporal filtering distortions, and 'open-loop' predictive coding of covered and uncovered regions. The 'open-loop' structure of TSB coding, however, affords two major advantages not enjoyed by MCP coding: simple optimal bit-allocation, non-existence of quantization error feedback. A new adaptive temporal subband (TSB) motion compensated predictive (MCP) coder is proposed. Hierarchical variable-sized block-matched regions with low predictive error are TSB coded, while poorly predicted regions are 'open-loop' MCP coded. Simulation results demonstrate that the adaptive coder substantially improves the temporal scalability of TSB coding, retains an advantageous 'open-loop' structure and provides comparable or superior PSNR to both MCP and TSB coding at MPEG-1 quality bitrates.

Indexing terms: video compression, motion compensation, temporal subband coding, scalable

1. INTRODUCTION

Multimedia communications methods designed for shared and distributed systems must be scalable. To robustly support many simultaneous independent video conferences and video retrieval sessions, scalable video compression is required. Scalable video compression permits extraction of coded video sequences at varying rates from one single compressed video-data bitstream, enabling variable network rate-control mechanisms. Two techniques exist for interframe coding for temporally scalable (variable frame rate) video: motion compensated predictive (MCP) coding, and temporal subband (TSB) coding [2], [3].

'Closed loop' predictive coding and subband coding have similar theoretical coding gains [4]. However, theory neglects the presence of quantization error feedback in a 'closed loop' coder, leading to reduced low-rate prediction efficiency, as well as the practical difficulty of calculating 'closed loop' optimal bit allocation. Despite this proviso, 'closed loop' MCP coders are still preferred to the 'open loop', which suffers from predictor divergence. Seemingly, subband coding may offer superior performance, as it does not suffer from predictor inefficiency. Indeed, subband and transform coding have been very successful for spatial coding. However, TSB coding suffers from two disadvantages: temporal filtering distortions, and 'open-loop' coding of covered and uncovered regions if regional motion compensation is employed ('Unlinked' regions in Figure 1 and 'inserted' and 'open-loop' pixels in Figure 2).

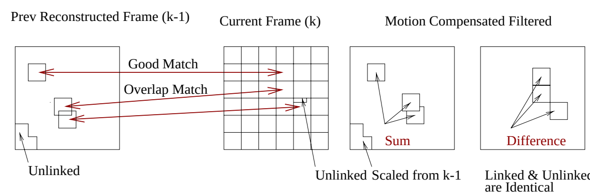

Figure 1: 3D Motion Compensated Subband Coding

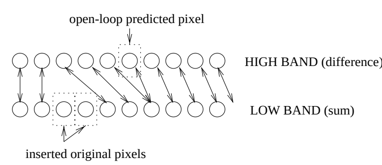

Figure 2: MC Temporal Analysis for TSB and TSB-MCP

For 2-tap Haar filter TSB coding, both MCP and TSB coding relate video frames in a similar fashion. However, while MCP coding transforms two related frames into one original (reference) frame and a second prediction error frame, 2-tap TSB coding produces a high frequency (Figure 1's 'Difference') and a low frequency frame (Figure 1's 'Sum') each containing components of the original two video frames. A 'telescoping' structure is a popular framework for relating frames of a video sequence to achieve temporal scalability at half, quarter, and sixteenth rates.

Lossless temporal scalability is not achievable with a TSB structure. To reduce the temporal rate high frequency frames must be dropped; however, in contrast to telescoping MCP, this may not be done without reducing the PSNR of the frames which are to be retained.

Adaptive TSB-MCP coding is motivated by the following rationale: temporal distortions may be reduced by 'open-loop' MCP coding regions with high prediction error, and coding gain is improved by utilizing TSB for other regions.

In this paper we demonstrate that an adaptive TSB-

Table 1: **PSNR Comparison of HVSBM temporally scalable hybrid TSB-MCP with TSB and MCP for test video sequences at 1.2 Mbps at full, half (Y@.5) , and quarter (Y@.25) temporal rates.**

Video Sequence	YUV band	MCP	rsMR MCP	TSB	adaptive TSB-MCP	adaptive MR TSB-MCP	open-loop MCP
Flower Garden	Y	26.11	26.47	26.68	26.55	26.70	26.25
	combo [1]	26.62	26.97	27.16	27.04	27.26	26.79
	Y @.5	26.83	27.12	27.06	27.49	27.64	
	Y @.25	26.73	27.94	27.36	27.96	28.03	
Table Tennis	Y	32.18	32.90	32.68	32.72	32.95	32.39
	combo	32.70	33.39	33.19	33.24	33.46	32.89
	Y @.5	32.94	33.57	32.26	32.89	33.39	
	Y @.25	33.68	34.24	32.53	33.08	33.57	
Football	Y	27.70	28.54	27.82	27.81	28.13	28.10
	combo	28.73	29.07	28.37	28.37	28.72	28.69
	Y @.5	28.99	29.64	27.61	28.63	29.19	
	Y @.25	30.42	30.90	28.11	29.37	29.90	
Mobile Calendar	Y	25.49	25.66	25.68	25.61	25.55	25.83
	combo	25.87	26.03	26.17	26.10	26.09	26.23
	Y @.5	26.08	26.23	26.15	26.54	26.52	
	Y @.25	26.52	26.63	26.29	26.74	26.74	

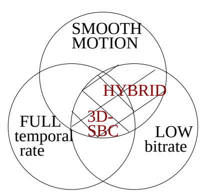

Figure 3: Expanded Scope of Applications of Adaptive MCP-TSB

MCP video coder is capable of producing a compressed video stream that is 'practically' scalable. That is, adaptive subband-predictive error frames may be dropped from the compressed video stream without significantly affecting the PSNR or visual quality of the reconstructed sub-rate video sequence. Most importantly, the adaptive framework provides competitive performance to 'closed-loop' MCP alone, while retaining the major advantage of TSB coding – simple optimal bit-allocation. As the Venn diagram in Figure 3 illustrates, applications requiring temporal scalability (such as videoconferencing) that are not feasible with 3D-SBC will be practical with an adaptive hybrid structure that improves scalability.

2. TEMPORAL SCALABILITY

In the following, input frames have Gaussian distribution with variance σ^2 and the variances of temporal subbands are $\sigma^2_{SB_k}$, $k = 0, \ldots, K-1$. The coding gain for TSB coding over intraframe coding for full-framerate is well-known as

$$G_{TSB/intra} = \frac{\sigma^2}{\left(\prod_{k=0}^{K-1} \sigma_k^2\right)^{\frac{1}{K}}}.$$

[1] Combo denotes the weighted average PSNR of all image YUV color bands.

A similar coding gain equation can be derived for $1/L$-framerate video, $L = 2^k, k \in Z^+$, extracted from TSB coding [1]

$$G_{TSB/intra} =$$

$$\frac{\sigma^2}{\frac{1}{\sqrt{L}}\left(\prod_{k=0}^{K_s-1} \sigma_k^2\right)^{\frac{1}{K_s}} + E\left[\left(F_{x,y,n} - \sum_{l=n}^{n+L-1} \frac{F_{x,y,l}}{L}\right)^2\right] 2^{2R_{avg}}}$$

where R_{avg} is the average rate used to code subband coefficients, subbands numbered $i = 0, \ldots, K_s - 1$ are those needed to synthesize the lower-framerate sequence, and $F_{x,y,z}$ is the original-frame pixel after motion-compensation that is located at spatial location (x, y) and time index n. The second term in the denominator is a function of bitrate and comes from temporal filtering distortions in the lower-framerate video. At very high rates total distortion in a frame asymptotically approaches the level of distortion caused strictly by temporal filtering. That is, temporally subsampled TSB cannot be a perfect reconstruction system unless the discarded subbands are zero energy (temporal filtering distortions do not exist).

Temporal filter distortions severely limit temporal scalability. For predictive coding at high rates, distortion of subrate video approaches zero. For MC-TSB coded video at less than full rate, the distortion approaches a lower bound determined strictly by the distortion due to temporal filtering. If the motion compensation of the video is imperfect, the temporal distortion may be large. The only method to improve the temporal scalability of the MC-TSB method is to reduce the magnitude of the introduced distortion.

While 'closed-loop' predictive coding uses compressed reconstructed frames for temporal prediction, 'open-loop' predictive coding employs original frames. The predictive frames employed by the encoder and decoder are identical for the 'closed-loop' method, but differ for the 'open-loop'

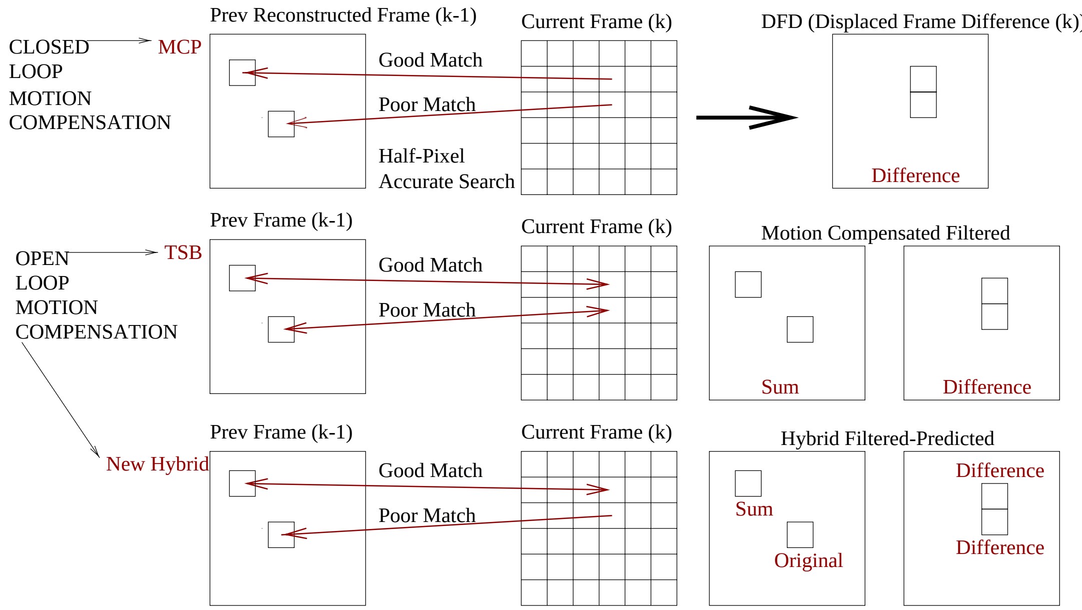

Figure 4: MCP, TSB, and Adaptive TSB-MCP

method (the original frames are not available at the decoder). This disparity causes 'open-loop' predictive coding to suffer from predictor divergence – each iteration increases the difference between the predictive frame in the encoder and that in the decoder.

If a sequence is to be locally motion compensated (an effective practical method for reducing video redundancy), not all image regions may be linked with a region in each other frame. In particular, uncovered and covered regions possess no corresponding region in some adjacent frames. For a MC-TSB coder, such regions must be 'open-loop' predictive coded – single pass 'closed-loop' coding is not possible.

To benefit from the advantages of MC-TSB coding, while maintaining high scalability, a method for limiting the maximum regional MC-TSB temporal distortion is proposed. At low and medium rates predictor divergence in a small number of regions is not a major impediment to achieving high performance. With the new adaptive method (at low and medium bitrates), the superior performance of MC-TSB for some sequences may be exploited without sacrificing temporal scalability.

3. METHOD

MC-TSB coding is usually combined with spatial subband coding (SBC) to unify the coding structure as completely subband based. This is the approach that is adopted for experimental results in this paper. The resulting structure is called three-dimensional subband coding (3D-SBC). As discussed above, MC-TSB is successful (superior to standard MCP) only at full temporal rate (with no temporal scalability) and at low-bitrate. Block motion-compensation artifacts (discontinuity at block-boundaries) decrease the efficiency of SBC for video interframes (versus block-transform coders). Combining MC-TSB with spatial SBC additionally limits success to smooth motion sequences that are more successfully motion-compensated than sequences with large abrupt motion and therefore contain less high-frequency energy at block-boundaries.

The cross-hatched area at the intersection of the Venn Diagram of Figure 3 indicates the current scope of applications that are appropriate for motion-compensated three-dimensional subband coding. MC-TSB is successful (superior to standard MCP) only at full temporal rate (with no temporal scalability) and at low-bitrate. Block motion-compensation artifacts (discontinuity at block-boundaries) decrease the efficiency of SBC for video interframes (versus block-transform coders). Combining MC-TSB with spatial SBC additionally limits success to smooth motion sequences that are more successfully motion-compensated than sequences with large abrupt motion and therefore contain less high-frequency energy at block-boundaries. The motivation for adaptive MCP-TSB is to expand the scope of applications to include applications that require temporal scalability (e.g., videoconferencing over the Internet or wireless channels). The expanded area labeled 'HYBRID' indicates that the desired structure is superior to standard MCP for smooth motion sequences at low bitrates.

The TSB structure of Choi and Woods [5] (Figure 2) is used because the 'open-loop' covered, uncovered and newly introduced regions are efficiently integrated with the temporal subbands. TSB blocks are HAAR filtered into high-pass and low-pass bands, prediction error from MCP blocks and pixels from covered/uncovered areas are scaled and placed in the high-pass temporal band, while unmatched pixels in the reference frame are scaled and placed in the low-pass band.

As the time delay between successive frames increases with each level of temporal decomposition, TSB distortions are increased due to inaccurate motion compensation. For this reason, TSB is used only for the first level of temporal processing. Further temporal decomposition employs MCP. The new adaptive encoder 'open-loop' codes all blocks with large mean absolute distance (MAD) from their predictors. If the MAD threshold is raised high enough the coder becomes 'open-loop' MCP, if it is lowered enough the coder becomes a somewhat scalable version of Choi and Wood's hierarchical variable size block matching motion compensated temporal subband coder [5].

In addition to the new TSB-MCP method, an enhancement is introduced – multiresolution block matching (MR BM) [6]. MR BM is a technique which improves the predictive capability of standard BM by adaptively choosing an optimal half-pixel interpolation filter for each motion compensated block. Although TSB cannot benefit from a multiresolution approach (a wide bandwidth half-pixel interpolation filter such as the MPEG 8-tap uniformly applied is best [7]), the 'open-loop' coded pixels which are a necessary part of regionally motion compensated TSB coding do benefit from MR compensation.

4. RESULTS

A MAD threshold of 15 gray levels is chosen for the adaptive TSB-MCP coder to satisfy the subjective criteria of reducing temporal distortion in the adaptive coder to a level consistent with the loss introduced through the coding process. Higher thresholds yield improved full rate PSNR at the expense of reduced sub-rate PSNR.

Although simple optimal bit allocation is a key advantage of adaptive TSB-MCP coding, all methods are compared with identical bit allocation to better enable comparison of relative coding gain performance of the methods. In practical applications, 'open-loop' MCP bit-allocation is often sufficient for 'closed-loop' MCP coding – the two converge at high bit rates, but for low bit rates 'closed-loop' MCP bit allocation may be far from optimal.

Ninety-six frames of the MPEG test sequences Flower Garden, Table Tennis, Football, and Mobile Calendar are coded at 1.2 Mbps, 30 fps, SIF resolution (352x240 pixels, 4:2:0 chrominance subsampled). The hierarchical variable sized block matching (HVSBM) uses a three level multiresolution quad-tree pyramid with largest block size 64x64 and smallest block size 4x4 pixels. Wavelet transform spatial coding follows temporal processing with a three level (10 subband octave) decomposition employing Daubechies' 9/7 bi-orthogonal wavelet pair, and is followed by adaptive arithmetic coding of quantized coefficient symbols, DPCM motion vectors, and the quad-tree map.

Five coding methods were examined experimentally, namely: motion compensated prediction (MCP) with wavelet prediction frame coding, reduced search multiresolution MCP (rsMR-MCP), temporal subband coding (TSB), adaptive TSB-MCP, and adaptive multiresolution TSB-MCP.

The primary observation in Table 1 is that while TSB offers a poorly scalable bit stream adaptive TSB offers equivalent scalability to MCP. For TSB, at half and quarter frame rate average PSNR relative to MCP is reduced .9 dB and 1.7 dB respectively, while for adaptive TSB average relative PSNRs are reduced less than .1 dB and .3 dB at both respective sub rates.

Also, while multiresolution processing improves MCP by .5 dB, it improves adaptive TSB-MCP by only .2 dB, – an effect due to the fact that only the MCP portion of TSB-MCP's blocks can take advantage of multiresolution prediction.

Overall, multiresolution adaptive TSB-MCP offers the best scalable compression for the Flower Garden and Mobile Calendar sequences. For the Table Tennis and high motion Football sequences rsMR-MCP is best. The same division is observed for the non-multiresolution variants of the TSB-MCP and MCP coders.

5. CONCLUSIONS

This paper presents a new category of inter-frame coding technique – adaptive TSB-MCP – designed to improve the temporal scalability of TSB coding while making its high coding gain available when possible.

Unfortunately, MC-TSB and MC-TSP-MCP both require 'two-pass' decoding. This added complexity in the decoder is a handicap for multicast, broadcast, database retrieval and other important asymmetric applications. Point-to-point communications (such as software-based videoconferencing) for which block-matching algorithms in the encoder dominate overall system complexity are, however, still viable applications for MC-TSB-MCP.

This new coder demonstrates temporally scalable performance which rivals and (for certain sequences) surpasses that of the ubiquitous MCP video coding structure at MPEG-I bit rates. It is advantageous for some sequences at low and medium bitrates, and may prove superior for advanced coders that offer superior motion compensation.

6. REFERENCES

[1] G.J. Conklin and S.S. Hemami, "Evaluation of temporally scalable video coding techniques," in *ICIP 97*, Santa Barbara, 1997, vol. II, pp. 61–64.

[2] D. Taubman and A. Zakhor, "Multirate 3-d subband coding of video," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 559–588, September 1994.

[3] J.R. Ohm, "Three dimensional subband coding with motion compensation," *IEEE Trans. Image Processing*, vol. 3, no. 9, pp. 559–571, Sept. 1994.

[4] J.W. Woods and S.D. O'Neal, "Subband coding of images," *IEEE Trans. Acoust. Speech Signal Process.*, pp. ASSP–34:1278–1288, Oct. 1996.

[5] S.J. Choi and J.W. Woods, "Three dimensional subband/wavelet coding of video with motion compensation," in *SPIE Visual Communications and Image Processing*, San Jose, CA, Feb. 1997.

[6] L. Winger and A.N. Venetsanopoulos, "Space-frequency motion model for subband/wavelet video coding," in *ICIP 98*, Chicago, October 1998.

[7] ISO/IEC CD 11172, *Coding of Moving Pictures and Associated Audio - For Digital Storage Media at up to 1.5 Mbits/sec*, Dec 1991.