PERIODIC PAN COMPENSATION FOR REDUCED COMPLEXITY VIDEO COMPRESSION†

Charles D. Creusere

Naval Air Warfare Center
China Lake, CA 93555
email: chuck@wavelet.chinalake.navy.mil

ABSTRACT

To reduce the complexity of a video encoder, we introduce a new approach to hybrid DPCM-tranform video compression in which pan compensation is performed outside the feedback loop. While the basic idea is conceptually similar to the pan compensation algorithm proposed by Taubman and Zakhor for their 3D subband coder, our method is different in that it continually tracks and updates the image in the feedback loop in the same way as a conventional hybrid coder. Using both residual energy and reconstruction error as metrics, we show that pan compensation implemented outside the feedback loop compares very favorably to similar compensation implemented within the conventional hybrid-transform framework. Furthermore, if the spatial coder used to compress the residual images outputs an embedded bit stream, then the complete system is spatially scaleable.

1. INTRODUCTION

In remote-sensing applications, severe constraints are placed on the design of the video encoder by weight, volume, and power limitations [1]. To further complicate matters, power and antenna constraints force the communications channel through which the video must be transmitted to have a narrow bandwidth, requiring that the compression algorithm operate efficiently at low bit rates. Unfortunately, achieving good rate-distortion performance generally requires a high complexity encoder which is capable of exploiting both inter- and intra-frame redundancies. For example, the most successful video compression technique, the hybrid DPCM-transform algorithm, has a far more complex encoder than decoder: exactly the opposite of what is required in this application. In remote-sensed video from aerial platforms (our primary area of interest), most of the motion is translation and, thus, it is generally sufficient to address only the problem of pan compensation. To this end, we introduce a new approach which has considerably lower encoder complexity than the conventional hybrid-transform algorithm while delivering equivalent performance.

This paper is organized as follows. Section 2 discusses conventional solutions to the hybrid video compression problem, pointing out their disadvantages with respect to the remote sensing application. In Section 3, periodic pan compensation is introduced and its tradeoffs are discussed. Section 4 presents comparative results while conclusions are presented in Section 5.

2. CONVENTIONAL MOTION COMPENSATION

2.1 Spatial Domain Compensation

The most common form of hybrid DPCM-transform video compression performs the motion compensation and differencing operations in the spatial domain (see Fig. 1 of [2]). This technique is used in all of the existing video compression standards including the motion picture experts group (MPEG) I and II standards for broadcast as well as the H.261 and H.263 standards for video teleconferencing. In these standards, the transform used is an 8×8 blocked discrete cosine transform (DCT). In terms of our remote sensing application, the complexity of this encoder is too high-- at least double that of the corresponding decoder. The problem of encoder complexity is further compounded for low bit rate applications because the low-complexity DCT is typically replaced with a higher complexity wavelet or subband decomposition. Doing this improves the rate-distortion performance of the video compression algorithm and eliminates many perceptually annoying artifacts (e.g. blocking), but it also greatly increases the complexity of the video encoder.

Assuming that F_k is the current frame in the video sequence, that an invertible transform is used, and that no quantization is applied, the frame reconstructed by the decoder is

$$\tilde{F}_k = F_k - MC\{F_{k-1}\} + MC\{F_{k-1}\} \quad (1)$$

for the hybrid DPCM-transform approach where MC is the motion compensation operation. Clearly, the frame reconstructed by the decoder is identical to that input to the

---
†Approved for public release, distribution unlimited

encoder, no matter what form of motion compensation is used. This property offers additional flexibility that can be used to improve the performance of the overall video compression system.

2.2 Transform Domain Motion Compensation

Numerous authors have proposed directly motion compensating the transform coefficients in a number of ways and for a variety of reasons. In [3] direct compensation of the coefficients was proposed so that spatial scalability could be introduced into the coded bit stream. The authors here noted that the maximally-decimated transforms typically used in coding applications are not spatially invariant and, therefore, that the compensation operations must be altered when applied in the transform domain. Specifically, to compensate the coefficients in a given subband, one must use coefficients from all of the subbands. Despite achieving performance equivalent to that of spatial compensation, the method of [3] is poorly suited to the needs of our remote sensing application because it also results in a significant increase in encoder complexity.

Other authors have instead proposed the direct implementation of motion compensation in the subband /transform domain. In this case, the compensation is separately performed in each band and interpolation is used to compensate for decimation in the analysis filter bank. Where direct comparisons have been performed, however, motion compensation in the subband domain has often proven inferior to equivalent compensation in the spatial domain [4].

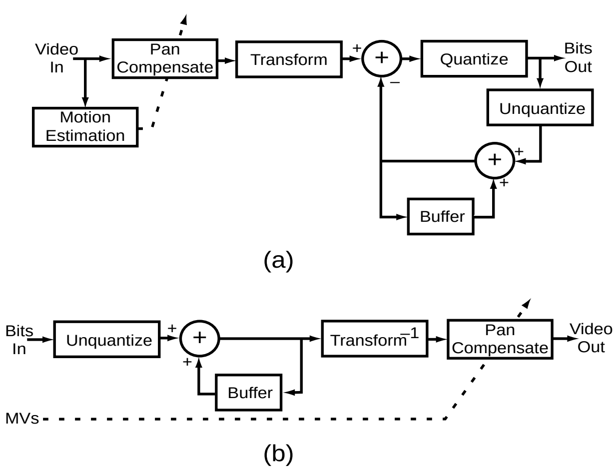

Figure 1: Hybrid video compression system using PPC. (a) is the encoder and (b) is the decoder.

3. PERIODIC PAN COMPENSATION (PPC)

Figure 1 illustrates the structure of the proposed video compression scheme. Note that the motion compensation is still performed in the spatial domain as with the conventional hybrid approach, but the differencing operation is now performed in the subband domain. Performing the differencing in the subband domain eliminates the need for the inverse transform in the feedback loop and, consequently, reduces the complexity of the video encoder by as much as 50% (depending on how the motion compensation and estimation are performed). By performing the actual motion compensation in the spatial domain, we overcome the problems inherent in the shift-varying nature of maximally decimated filter banks and transforms.

Unfortunately, there is a major limitation to the motion compensation structure of Fig. 1. To see this, we can express the frame reconstructed by the decoder as

$$\tilde{F}_k = MC^{-1}\left\{MC\left\{F_k\right\} - F_{k-1} + F_{k-1}\right\}, \qquad (2)$$

assuming again, as in (1), that no quantization is performed and that the transform is perfect reconstruction. Unlike the conventional scheme, (2) requires that the motion compensation operation be invertible. A motion compensated video compression algorithm with a structure similar to that of Fig. 1 has been proposed by Taubman and Zakhor [5]. Their coder uses a 3D subband decomposition and compensates for motion only over the support interval of the temporal filter bank, modifying the filters at the interval boundaries to achieve the invertibility required by (2). In this work, we do not assume that the motion compensation is performed over a finite set of frames; instead, we require that the system continually track and update the motion estimate, regardless of the number of differentially predicted frames between each I-frame (e.g., each frame coded using only information contained within it).

The video compression system illustrated in Fig. 1 can compensate for translational or panning motion if pixels which would normally move out of the input frame during the compensation process are instead wrapped around to its other side. If F(x,y) is the X×Y image to be compensated using the integer motion vector $(\Delta m, \Delta n)$, then periodic pan compensation is given by

$$PPC\left\{F(x,y)\right\} = F\begin{pmatrix}(x-\Delta m)\bmod(X), \\ (y-\Delta n)\bmod(Y)\end{pmatrix} \qquad (3)$$

where 'mod' is the standard modulo operation.

By making a few reasonable assumptions about the statistics of the video sequence, we can compare conventional and periodic pan compensation. Because pan compensation is implemented separably in the horizontal and vertical directions, it suffices here to consider only a 1-dimensional (1D) system. In addition, since the adverse effects of PPC are not cumulative, we need only consider the compensation of a single frame. Assume that the input $\mathbf{x}$ is a random vector in $\mathbf{R}^N$ and that motion compensation is performed by a matrix multiplication, i.e., $\tilde{\mathbf{x}} = \mathbf{V} \cdot \mathbf{x}$. We further assume that the mean of this vector is given by

$$\text{mean} = E\{\mathbf{x}\} = \mathbf{0} \tag{4}$$

and the covariance by

$$\text{cov}_{j,k} = E\{x(j) \cdot x(k)\} = \alpha^{|j-k|} \cdot \sigma^2 \tag{5}$$

where $E\{\bullet\}$ is the expectation operator and $0 \le \alpha < 1$. If the new input is truly just a shifted version of the previous input, then $\hat{\mathbf{x}}$ (the previous input which we are trying to compensate $\mathbf{x}$ to match) must also satisfy (4) and (5) with the additional condition on the cross correlation that

$$E\{x(j) \cdot \hat{x}(k)\} = \alpha^{|j-k+\Delta n|} \cdot \sigma^2 \tag{6}$$

when the input vector is shifted by Δn with respect to $\hat{\mathbf{x}}$. Equation (6) states that the statistical correlation between elements in the signal decreases as the distance between them increases. This common statistical image model is known to hold well over limited regions of an image, but we apply it broadly in order to characterize the worst case performance of PPC relative to conventional pan compensation.

When using the conventional hybrid transform approach, one need not wrap pixels around the edges of the image as is done in (3) to satisfy (2). Rather, one can instead use pixel replication to fill in the areas of the image uncovered by motion; given (6), this is the optimal course of action. In general, the residual vector after motion compensation is given by

$$\varepsilon = \tilde{\mathbf{x}} - \hat{\mathbf{x}} \tag{7}$$

and the corresponding expected squared error by

$$E\{f\} = E\{\varepsilon^t \cdot \varepsilon\} = E\left\{(\tilde{\mathbf{x}} - \hat{\mathbf{x}})^t \cdot (\tilde{\mathbf{x}} - \hat{\mathbf{x}})\right\} \tag{8}$$

where $(\bullet)^t$ is the transpose operation and f is simply the sum of the squared errors. Starting from (8) and applying (4) through (6), it can be shown that

$$E\{f_C(\alpha, \Delta n)\} = 2\sigma^2 \cdot \left(|\Delta n| - \frac{\alpha}{1-\alpha}\left(1 - \alpha^{|\Delta n|}\right)\right) \tag{9}$$

where f_C is the error in the case of conventional pan compensation.

If periodic pan compensation (as defined by (3)) is used in the context of Fig. 1, the expected squared error is given by

$$E\{f_{PPC}(\alpha, \Delta n)\} = 2\sigma^2 |\Delta n| \cdot \left(1 - \alpha^N\right). \tag{10}$$

The difference between (9) and (10) represents the loss of efficiency incurred by using PPC over conventional compensation and is given approximately by

$$\text{diff} = \frac{2\sigma^2 \cdot \alpha}{1-\alpha}\left\{1 - \alpha^{|\Delta n|}\right\} \tag{11}$$

for large N and realistic values of α. To characterize the performance degradation graphically, we form the ratio

$$\text{ratio} = \frac{N \cdot \sigma^2 - \text{diff}}{N \cdot \sigma^2} = 1 - \frac{2\alpha}{N \cdot (1-\alpha)}\left(1 - \alpha^{|\Delta n|}\right) \tag{12}$$

and plot it in Fig. 2 for N = 512 (the length of one line in a typical image) and a few values of α. From the figure, we note that the penalty paid for using PPC (i.e., the reduction from 1.0) is not truly severe even for large values of Δn and α, leading us to expect that the method will perform well in practice. The performance of PPC relative to conventional pan compensation improves as the regional correlation in the statistical image model decreases because the conventional algorithm's flexibility in filling the part of the frame uncovered by motion becomes less significant.

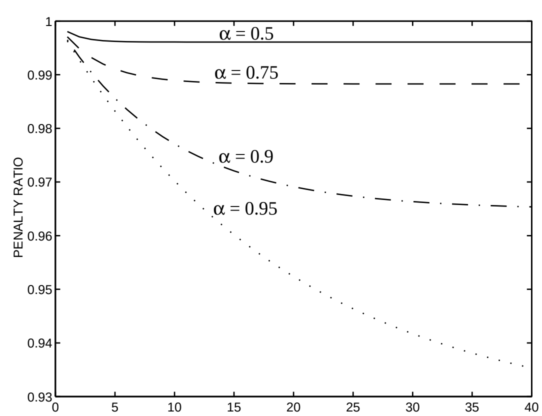

Figure 2: Difference in the expected error between conventional and periodic pan compensation for a variety of α.

4. RESULTS

The comparative results are summarized in Fig. 3 where the input is a 30 frame aerial sequence of a power plant having 512×512 pixels per frame. In the figure, ‘ENERGY’ refers to the sum of the squared pixel values of the residual image divided by the total number of pixels and ‘MSE’ refers to the mean squared error between the original image and its reconstruction. We use here a 5-level decomposition based on the 9/7 biorthogonal wavelet, and we code the resulting coefficients using Shapiro’s embedded zerotree wavelet (EZW) algorithm [6]. The first frame of the sequence is intra-frame compressed to 0.2 bits/pixel (bpp) while the remaining frames are differentially compressed to 0.1 bpp. Motion vectors are computed by using correlation between centered 128×128 patches in the current and previous frames. Examining Fig. 3a, we note that the residual energy of PPC using circular convolution is essentially identical to that of the conventional pan compensation (using either circular convolution or symmetric extension to cancel filter startup transients). Interestingly, PPC with symmetric extension actually has the smallest reconstruction error (see Fig. 3b) despite its higher residual energy. For this video sequence, it is clear that using periodic pan compensation within the framework of Fig. 1 does not adversely effect on the objective performance of the compression system. Furthermore, we have also found that the decoded and reconstructed frames produced by both proposed and conventional video compression schemes have very similar perceptual quality.

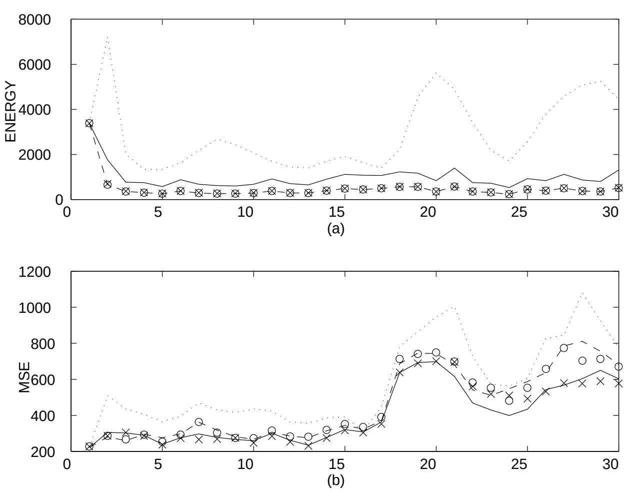

Figure 3: (a) Uncoded difference energy and (b) MSE reconstruction error versus frame number. Dotted is uncompensated; solid and dashed are symmetrically extended periodic and non-periodic compensation; and X’s and O’s are periodic and non-periodic panning using circular convolution.

5. CONCLUSION

We have presented here a new approach to motion compensation which reduces the complexity of the video encoder by eliminating the inverse transform operation. If integer pixel compensation is desired and the motion vectors are generated by sources external to the encoder (i.e., inertial sensors and gimbal angles), then the complexity reduction achieved by using the new out-of-loop compensation scheme over conventional hybrid DPCM is 50%. In addition, comparisons made using a representative video sequence show that the new approach generates rate-distortion and residual energy results comparable to those of the conventional method for a variety of specific implementations. Finally, we note that the structure of Fig. 1 is scaleable in powers of 2 since the differencing operation is performed in the transform domain and the decoder can thus process only those coefficients corresponding to a desired resolution without losing its synchronization to the encoder.

REFERENCES

[1] C.D. Creusere and Gary Hewer, “Digital video compression for weapons control and bomb damage indication,” *AGARD Conference Proceedings 576*, Chapter 16, Sept. 1995.

[2] H. Gharavi, “Subband coding algorithms for video applications: videophone to HDTV-conferencing,” *IEEE Trans. on Circuits and Systems for Video Tech*., Vol. 1, No. 2, June 1991, pp. 174-183.

[3] C. Cafforio, et. al., “Motion compensation and multiresolutional coding,” *Signal Processing: Image Communications* 6 (1994), pp. 123-142.

[4] K. Sawada and T. Kinoshita, “Subband-based scalable coding schemes with motion compensated prediction,” *Proc. of the SPIE*, VCIP 1995, Vol 2501, pp. 1470-1477.

[5] D. Taubman and A. Zakhor, “Multirate 3-D subband coding of video,” *IEEE Trans. on Image Proc*., Vol. 3, No. 5, Sept. 1994, pp. 572-588.

[6] J. Shapiro, “Embedded image coding using zerotrees of wavelet coefficients,” *IEEE Trans. on Signal Proc.*, Vol. 41, No. 12, Dec. 1993, pp. 3445-3462.