ON MOTION COMPENSATION OF WAVELET COEFFICIENTS†

Hiroshi Ito‡ and Nariman Farvardin

Electrical Engineering Department and Institute for Systems Research
University of Maryland
College Park, MD 20742, USA

ABSTRACT

New intra- and inter-frame coding techniques for coding a wavelet decomposed video signal are proposed. These methods are based on the required accuracy of motion compensation for each of the different frequency subbands. For each subband, our methods prohibit inter-frame coding and switch to intra-frame coding when the estimated motion vector is not sufficiently accurate. The simple switching mechanism adopted results in a low overhead for coding the adaptation parameters while keeping the loss in coding gain negligible, compared to an unconstrained adaptive strategy in which intra- and inter-frame coding can be switched independently in each subband. The required accuracy of motion compensation is derived analytically for equally distributed pass-band signals and used throughout the paper. Simulation results for practical video sequences are presented.

1. INTRODUTION

Over the past few years there has been much activity in image coding using the discrete wavelet transform [1, 2, 3, 4]. In coding of video sequences, the inter-frame correlation is generally exploited by temporal prediction with motion compensation [5]. However, it is reported that in transform coding situations, motion compensation does not work well for the high frequency coefficients [6, 7]. In this work, we compute the required accuracy of motion compensation for each wavelet coefficient and based on this result propose adaptive intra- and inter-frame coding strategies.

2. SYSTEM DESCRIPTION

The input video signal is first decomposed, on a frame-by-frame basis, into ten subbands, LL, H_i, V_i and D_i, $i = 1,2,3$, as shown in Fig. 1. Subbands H_i, V_i and D_i are grouped together and denoted B_i, $i = 1,2,3$ and LL is denoted B_0, i.e.,

$$B_i = \begin{cases} LL & \text{if } i = 0 \\ H_i \cup V_i \cup D_i & \text{otherwise.} \end{cases} \quad (1)$$

With this frequency splitting, the system offers four different levels of spatial resolution. The overall structure of the encoding system is shown in Fig. 2. Each subband is coded by a frame prediction coder followed by a quantizer and possibly a variable-length coder. In this paper, we do not consider the quantization problem and hence quantization errors are assumed to be zero. The frame-to-frame motion is estimated by a block matching technique using the current frame and the reconstructed version of the previous frame. Note that, since we assumed no quantization errors, the previous frame is exactly the frame located before the current frame in the original video sequence. The same motion vectors are then used for motion compensation in all subbands. The previous frame is stored separately in a frame memory. For motion compensation, each subband is expanded to the original size using the wavelet synthesis filters, displaced by the amount of estimated motion and subsampled again. This predicted frame is subtracted from the current frame and prediction errors are obtained for each subband. The variance of the prediction error is compared to that of the original wavelet coefficients on a block-by-block basis and the one with a smaller variance is selected for transmission. This adaptation is the main contribution of this paper and will be described in more detail in subsequent sections.

3. ESTIMATION OF ACCURACY

For simplicity, consider a one-dimensional, continuous-time bandpass signal $g(t)$ with a constant spectrum in the frequency band from w_1 to w_2 (rad/sec). The frame difference with a motion estimation error of aT

‡On leave from Mitsubishi Electric Co.

†This work was supported in part by a grant from Mitsubishi Electric Co. and in part by a grant from the National Science Foundation NSFD CD 8803012.

0-7803-2431-5/95 $4.00 © 1995 IEEE

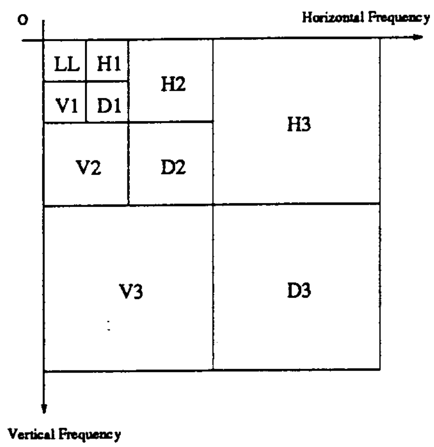

Figure 1: Subband Decomposition

(T being the sampling interval) is given by

$$d(t) = g(t) - g(t - aT). \qquad (2)$$

Taking the Fourier transform and comparing the power of the signals $g(t)$ and $d(t)$, it is easy to see that the prediction gain G_p is given by

$$\frac{1}{G_p} = 2 - \frac{2\{\sin(aw_1T) - \sin(aw_2T)\}}{aT(w_1 - w_2)}. \qquad (3)$$

Table 1 shows the prediction gains of actual video sequences obtained from a computer simulation. Here, we gave specific displacements to the original frames and computed the resulting prediction gains. The results show a good correspondence between the values computed from Equation(3) and those obtained from displaced video sequences, except for the LL band, in which the flat spectrum assumption is not valid. It is also seen that with the half pixel displacement, the prediction gain for higher frequency subbands is larger in simulations than what the analytical values suggest. This is due to the low-pass filtering effect of interpolation. We have obtained similar results for vertical displacements.

From Equation (3), we see that for the prediction gain to be greater than 1 the following condition needs to be satisfied

$$\frac{\sin(aw_1T) - \sin(aw_2T)}{aT(w_1 - w_2)} > 1/2. \qquad (4)$$

In the two-dimensional case, the above inequality gives the required accuracy in horizontal and vertical directions, respectively. Table 2 shows the maximum horizontal values of the parameter a which satisfy Equation (4) for several subbands splitted as in Fig. 1. Other values are calculated in the same way. As shown in Table 2, a more accurate motion compensation is required for higher subbands, or equivalently, with the same accuracy the prediction gain reduces as the frequency gets higher.

Table 3 shows simulation results of motion compensated prediction gains obtained with integer pixel accuracy for the "Miss America" sequence. A high prediction gain is obtained in the low frequency band implying that motion is tractable in most blocks of this sequence. However, the gain decreases monotonically for higher frequency bands until there is little gain for the highest frequency band. We have observed that in

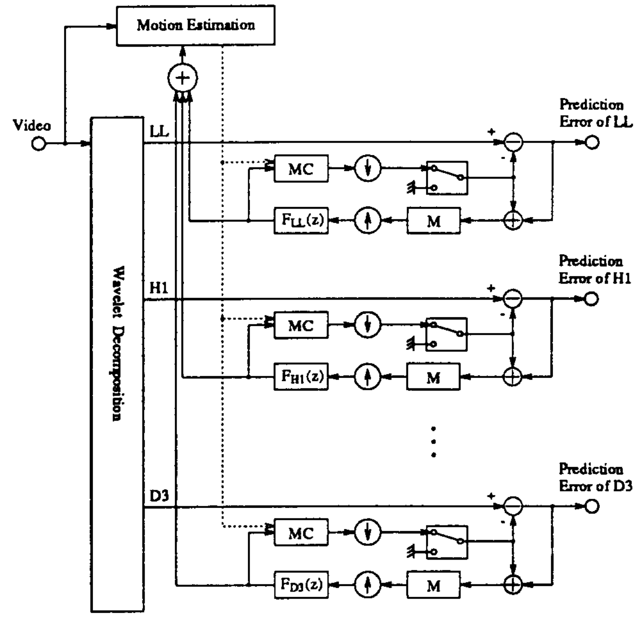

Figure 2: System structure(M:Frame store, MC:Motion compensator).

D	Sequence	Band			
		LL	H_1	H_2	H_3
0.5	Analytical	18.92	10.50	4.59	-0.98
	M.America	29.50	10.91	5.57	0.92
	Salesman	26.27	9.33	4.88	1.58
1.0	Analytical	12.92	4.59	-0.98	-5.15
	M.America	24.16	5.20	0.10	-4.52
	Salesman	20.64	3.45	-1.10	-4.39

Table 1: Relationship of horizontal displacement and inter-frame prediction gains (averaged over frames 74-81 of "Miss America" and frames 106-113 of "Salesman", D: displacement in pixels).

	LL	H_1	H_2	H_3
a	4.83	1.76	0.88	0.44

Table 2: Required accuracy (in pixels) of motion compensation for horizontal subbands.

most cases the estimation error is within 1/2 pixels. We have also seen that this indicates that the motion estimation has been preformed fairly well.

	B_0	B_1	B_2	B_3
G_p	29.08	12.26	6.65	0.70

Table 3: Prediction gain (in dB) of each subband ("Miss America", with integer pixel motion compensation).

4. ADAPTATION STRATEGIES

Based on the discussion in the previous section, one can conclude that higher frequency subbands need a more accurate motion compensation. However, in practice, widely used block matching techniques provide insufficient accuracy when motion is complicated. In many video coding situations, block-based adaptation between intra- and inter-frame codings is necessary. However, it is clear that such a method would not work well if the same criteria for adaptation are used in all subbands. In what follows, we propose two adaptive algorithms for switching between intra- and inter-frame coding on a block-by-block basis.

Top-Down Algorithm

Step 1: For the lowest frequency subband, choose between intra- and inter-frame coding based on the variance of the original signal and the prediction error.

Step 2: If intra-frame coding is chosen in B_{i-1}, then choose intra-frame coding again in B_i. Otherwise, choose between intra- and inter-frame coding based on the variance of the original signal and the prediction error.

Step 3: Repeat Step 2 until all subbands are processed.

The restriction in Step 2 reduces the overhead for intra/inter-frame adaptation with little decrease in coding gain compared to the unconstrained adaptive switching. It also improves the subjective picture quality by prohibiting motion compensation when the estimates of motion vectors are inaccurate. However, in this algorithm, a decision error in a low frequency subband propagates to subsequent subbands and sometimes increases the variance of the signal to be transmitted.

An alternative to the top-down approach is a full-search algorithm where a set of switching parameters which minimizes the resulting signal variance is selected. Based on the results of the previous section, however, we place a restriction on the switching parameters so that intra-frame coding is selected in the n highest subbands, where n varies from 0 to M, the number of subbands.

Full-Search Algorithm

Step 1: Set all subbands to inter-frame coding and compute the variance of the resulting prediction errors; E_0.

Step 2: Set $n = 1$; set the n highest frequency subbands to intra-frame coding and compute the variance of the signal to be transmitted; E_n.

Step 3: Set $n = n + 1$; if $n \leq M$ go to Step 2. Else; pick $n^* = \arg\min E_n$. Use intra-frame coding for the n^* highest frequency bands and inter-frame coding for all others.

5. SIMULATION RESULTS

Table 4 lists the coding gains when the above switching method is used for "Miss America" with integer pixel accuracy for motion compensation. For simplicity, the LL band is set to choose inter-frame coding all the time. For comparison, coding gains of (non-adaptive) inter-frame prediction and unconstrained adaptive switching are included. In both the top-down and the full-search algorithms, adaptivity leads to some improvement. In the top-down algorithm, the improvement is somewhat limited in B_3. This is due to the fact that B_3 suffers most from an inaccurate decision of using intra-frame coding in the lower subbands, as discussed in the previous section. It can be seen that this is avoided in the full-search algorithm, in which the coding gain of the highest subband is very close to that of the unconstrained adaptive case. Table 5 shows the results when motion compensation is performed using half-pixel accuracy. Here, we can see that the coding gain in the highest frequency band gets worse by the adaptation of the top-down algorithm. Still the full-search algorithm shows a good result. Similar results for the "Salesman" sequence are shown in Tables 6 and 7.

	B_0	B_1	B_2	B_3
Top down	29.66	12.69	6.87	0.77
Full search	29.66	12.66	6.90	1.13
Non adaptive	29.66	12.54	6.73	0.16
Unconstrained	29.66	12.69	6.92	1.13

Table 4: Adaptive prediction gain (in dB) for each subband (Miss America, integer pel accuracy).

	B_0	B_1	B_2	B_3
Top down	31.59	15.10	8.69	1.29
Full search	31.59	15.04	8.65	1.70
Non adaptive	31.59	14.97	8.61	1.59
Unconstrained	31.59	15.10	8.69	1.70

Table 5: Adaptive prediction gain (in dB) for each subband (Miss America, half pel accuracy).

6. CONCLUSIONS

We have investigated the required accuracy of motion estimation for subband images and found that for higher frequency subbands the estimation error should be less than half a pixel. This implies that motion estimation tends to fail more easily for higher frequency subbands. Based on this investigation, we have proposed effective motion compensation strategies where an intra/inter frame switching is performed hierarchically. Computer simulations show that our proposed techniques perform reasonably well. The sequences we have used so far ("Miss America" and "Salesman") are relatively easy for motion compensation purposes. In more complicated scenes, more intra-frame coding will be chosen by our algorithm for higher frequency bands and the advantages of the proposed scheme will be more pronounced. Work on extending our results to other sequences and on incorporating the proposed adaptation mechanism in a low-rate video coding scheme designed for operation in a wireless environment is currently underway.

	B_0	B_1	B_2	B_3
Top down	21.95	12.15	11.07	6.09
Full search	21.95	12.13	11.14	6.11
Non adaptive	21.95	11.71	10.82	5.83
Unconstrained	21.95	12.15	11.16	6.12

Table 6: Adaptive prediction gain (in dB) for each subband (Salesman, integer pel accuracy).

	B_0	B_1	B_2	B_3
Top down	22.38	12.76	11.66	6.59
Full search	22.38	12.72	11.77	6.62
Non adaptive	22.38	12.36	11.50	6.54
Unconstrained	22.38	12.76	11.79	6.62

Table 7: Adaptive prediction gain (in dB) for each subband (Salesman, half pel accuracy).

7. REFERENCES

[1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image Coding Using Wavelet Transform", *IEEE Trans. Image Process.,* vol. 1, pp. 205-220, Apr. 1992.

[2] J. W. Woods and S. D. O'Neil, "Subband Coding of Images", *IEEE Trans. ASSP,* vol. ASSP-34, pp. 1278-1288, Oct. 1986.

[3] P. Sriram and M. W. Marcellin, "Image Coding Using Wavelet Transforms and Entropy-Constrained Trellis Coded Quantization", *to appear in IEEE Trans. Image Proc.,* 1994.

[4] A. Said and W. A. Pearlman, "Image Compression Using the Spatial-Orientation Tree", *Proc. IEEE Int. Symp. Circuit and Systems* , Chicago, IL, pp. 279-282, May 1993.

[5] K. Tsunashima, J. B. Stampleman and V. M. Bove, "A Scalable Motion-Compensated Subband Image Coder", *IEEE Trans. Commun.*, vol. 42, pp. 1894-1901, Feb.-Apr. 1994.

[6] H. Yamaguchi, "Adaptive DCT Coding of Video Signal", *IEEE Trans. Commun.*, vol. 41, pp. 1534-1543, Oct. 1993.

[7] B. Girod, "The Efficiency of Motion-Compensating Prediction for Hybrid Coding of Video Sequences", *IEEE Journal of Selected Areas in Commun.*, vol. SAC-5, pp. 1140-1154, Aug. 1987.