A NEW TECHNIQUE FOR MOTION ESTIMATION AND COMPENSATION OF THE WAVELET DETAIL IMAGES

G. Van der Auwera, A. Munteanu, G. Lafruit[*] and J. Cornelis

Vrije Universiteit Brussel
ETRO/IRIS Research Group
Pleinlaan 2, 1050 Brussels, Belgium
[*]IMEC - Leuven, Belgium
e-mail: gvdauwer@etro.vub.ac.be

ABSTRACT

This work proposes a new block based motion estimation and compensation technique applied on the detail images of the wavelet pyramidal decomposition. The algorithm implements two block matching criteria, namely the absolute difference (*AD*) and the absolute sum (*AS*). To assess the coding performance of this method, we have implemented a software simulation of a wavelet video encoder based on our square partitioning coder. Coding results indicate a considerable image quality gain, expressed by PSNR values, of up to 3.4 dB compared to intra wavelet coding for the same bit rate. The quality gain, obtained by using two matching criteria (*AS* and *AD*) instead of one (*AD*), varies between 0.3 and 0.5 dB.

1 INTRODUCTION

The usage of wavelets to compress video sequences is subject to research. The possible benefits are an improved image quality for very low bit rate applications with respect to current standards (H.263) and the feasibility of progressive transmission for variable channel bandwidths.

In this paper, we briefly review a wavelet based video encoder architecture performing motion estimation and compensation in the detail images of the wavelet transform. For this encoder we propose a new block based motion estimation algorithm implementing two block matching criteria, namely the absolute sum (*AS*) and the absolute difference (*AD*). The use of two matching criteria is motivated by our work reported in [8] and which we summarize in this paper.

We also present coding results by applying our square partitioning coder (*SQP*) [6]. We compare the reconstructed image quality for the “Mobile & Calendar” sequence to that of the MPEG-4 Verification Model (*VM*) [3]. We conclude this paper by presenting a PSNR curve surpassing that of *VM* for an identical bit rate.

2 WAVELET BASED VIDEO CODEC

A straightforward approach to build a wavelet based video codec, is to replace the DCT in a classical video coder by the discrete wavelet transform [2][5]. If the discrete wavelet transform (*DWT*) is removed from the prediction loop, the video encoder architecture depicted in Fig. 1 is obtained. Both motion estimation (*ME*) and compensation (*MC*) are performed in the wavelet domain, i.e. in the average image of the highest decomposition level and in the detail images. This is feasible since the wavelet transformed image contains not only frequency information but also spatial information, which is not the case for the DCT.

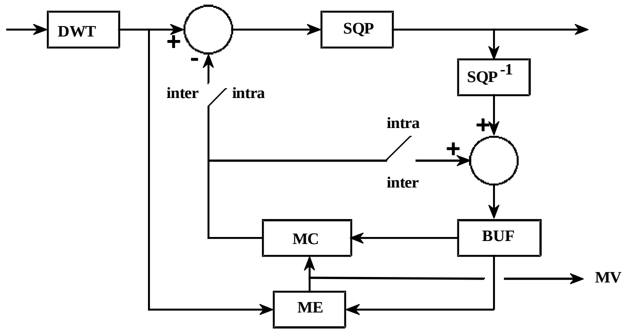

Figure 1. Wavelet based video encoder without inverse discrete wavelet transform.

However, difficulties are encountered with this approach, because in general the discrete wavelet transform is not shift invariant [1], due to the subsampled nature of the transform. This implies that shifts in the spatial domain do not just produce shifts in the wavelet domain subimages, but change the values of the coefficients in these subimages as well. However, there is an exception if the shifts in the spatial domain are multiples of the sampling period. A dyadic wavelet transform is completely shift invariant if the spatial domain shift has the form $k \cdot 2^J$, $k \in Z$, where J denotes the number of decomposition levels.

Some motion estimation methods have already been introduced in [4][9]. They perform a hierarchical motion estimation in the wavelet detail images by using the mean absolute difference (*MAE*), or the mean square difference (*MSE*) as block matching criterion. However, since spatial shifts produce ambiguous effects in the wavelet image, we conclude that new methods are required for motion estimation and compensation in the wavelet domain.

3 ALGORITHM DESCRIPTION

In this section we introduce a new motion estimation algorithm in the wavelet detail images. The new method uses two block matching criteria, namely *AD* and *AS*, given by:

$$AD = \sum_{i,j} \left| o(i,j) - c(i,j) \right| \text{ and } AS = \sum_{i,j} \left| o(i,j) + c(i,j) \right| .$$

$o(i,j)$ is the original block containing wavelet coefficients and $c(i,j)$ is a candidate block for prediction. The best candidate minimizes *AD* or *AS*. In section 4 we summarize the results of [8] which lead to the conclusion that *AS* often achieves a lower prediction error than *AD*. In section 5 we will compare the coding efficiency of this algorithm to the method implementing only *AD* as matching criterion, which is described in the following paragraph.

3.1 FS using Absolute Difference

This method performs full-search motion estimation on every level of the wavelet decomposition by using *AD* as error criterion. In our simulations, we use a 3 levels wavelet decomposition, so the full-search motion estimation is performed in the four subimages of level 3 and in the six subimages of levels 2 and 1. To define the block sizes in the detail images we use two different approaches. In the first one we impose the same block size in the detail images, while in the second we use dyadic block sizes containing $2^{c-j} \times 2^{c-j}$ coefficients, where j denotes the decomposition level and c is a constant. We identify this algorithm as *FS-AD* (full-search using *AD*) in the section reporting the coding results.

3.2 FS using Absolute Sum and Difference

We propose a new algorithm that performs full-search motion estimation on every level of the wavelet decomposition and implements two matching criteria for finding the best block, namely *AS* and *AD*. The block sizes on every level are specified as for the *FS-AD* method. In the average image we use only *AD* as matching criterion due to its lowpass nature.

In the *FS-AD* method, the motion vector is determined by the position of the block in the reference image that minimizes *AD*. If we also calculate *AS* for every search position in the reference image, then it is possible that the minimum obtained with the *AS* criterion is smaller than the minimum given by the *AD* criterion. We deduce that this method yields a smaller prediction error than the *FS-AD* method. We refer to this algorithm as the *FS-AS/AD* method (full-search using *AS* and *AD*).

3.3 Implementation Details of FS-AS/AD

One bit for each predicted block has to be recorded as side information to distinguish between the motion vectors determined by *AD*, respectively *AS*. In the encoder (Fig. 1) the motion compensation can use this information to change the signs of the predicted block coefficients for which the *AS* criterion has been used, so that when the predicted wavelet image is subtracted from the original image these blocks are summed.

In [7] we describe the arithmetic complexity of the *FS-AS/AD* and *FS-AD* method. It follows that *FS-AS/AD* takes twice the number of arithmetic operations of the *FS-AD* method, because it makes use of two block matching criteria in parallel. The arithmetic complexity determines the required hardware, but it is not the only factor to take into consideration. If one also considers energy dissipation, then the memory transfers will be the dominant factor. In [7] we conclude that the extra arithmetic complexity of the *FS-AS/AD* method with respect to *FS-AD* is negligible if one considers energy dissipation, since the number of required memory transfers does not significantly increase.

4 1D-STEP SHIFT COMPENSATION

In [8] we show that the prediction error of the detail images can be reduced if one considers both summing and subtracting the original and the candidate blocks. We summarize the analysis in this section.

The detail images contain high frequency information which corresponds mainly to edges in the spatial domain. The analysis [8] is restricted to the one-dimensional case, and we model an arbitrary edge by a step profile $x(n)$. Suppose $x_g(n)$ is the highpass component obtained from a one level wavelet analysis of $x(n)$ by the biorthogonal filter $g(n)$. Denote by $x_g(n-s)$ the signal obtained by shifting with s positions the wavelet component $x_g(n)$, and by $y(n)$ the signal obtained by shifting with k positions the original signal $x(n)$: $y(n) = x(n-k)$. The highpass component of a one level wavelet analysis of $y(n)$ is $y_g(n)$. If k is even, it is proven in [1] that the one level wavelet transform is shift invariant, therefore we obtain a zero prediction error if we subtract the original samples $y_g(n)$ and the predicted samples $x_g(n-k/2)$. Conversely, for odd shifts of $x(n)$, the absolute sum (*AS*) between the predicted samples $x_g(n-s)$ and the original samples $y_g(n)$ is lower than the absolute difference (*AD*). In [8] we show this for the particular case $k=1$ and $s=1$. The values of *AD* and *AS* are evaluated for different biorthogonal filters $g(n)$ in Table 1. As we note, the absolute sum is smaller than the absolute difference for all the considered filters. We observe also that *AS* is zero for the first two filters. Hence, a zero prediction error can be obtained.

Filters	AS	AD
Biorthogonal (2.4)	0.0000	0.3535
Biorthogonal (2.8)	0.0000	0.3535
Biorthogonal (3.9)	0.3535	0.4419
Biorthogonal (5.5)	0.1772	0.5459
Biorthogonal (6.8)	0.1315	0.4342
Biorthogonal (9.7)	0.0883	0.4349

Table 1. The values of the absolute sum and absolute difference for different biorthogonal filters $(s=1)$.

5 WAVELET VIDEO CODING

5.1 Description of the Wavelet Video Encoder

To assess the performance of the *FS-AS/AD* method, we have implemented a software simulation of the wavelet encoder architecture depicted in Fig. 1. To code the intra images (*I*) as well as the inter images (*P*), we use our square partitioning coder [6] followed by an arithmetic coder. We have chosen the biorthogonal (9,7) wavelet filters to generate a 3 levels pyramidal image structure for the motion estimation process. This choice is inspired by the fact that these filters in general provide the best coding results for photographic images. Moreover, in Table 1 we have shown that for an odd shift of the step function the prediction error obtained by the *AS* criterion is very low.

5.2 Square Partitioning Coder

The square partitioning coder [6] applies a wavelet based image coding technique that exploits the dependencies within the wavelet subbands. Successive approximation quantization (*SAQ*) is applied to provide a multiprecision representation of the coefficients and to facilitate the embedded coding.

The significance of the wavelet coefficients with respect to a monotonically decreasing series of thresholds, is determined by using *SAQ*, and is indicated in binary maps called significance maps. For each threshold, the corresponding significance map is encoded efficiently using a hierarchical tree structure of squares that group the insignificant coefficients in blocks of variable width.

With this coding scheme, a prioritisation protocol is implemented, in which the ordering of importance is determined, by the precision, magnitude, scale and spatial location of the wavelet coefficients. The bit-stream that results is completely embedded, so that all the versions of the encoded image at lower

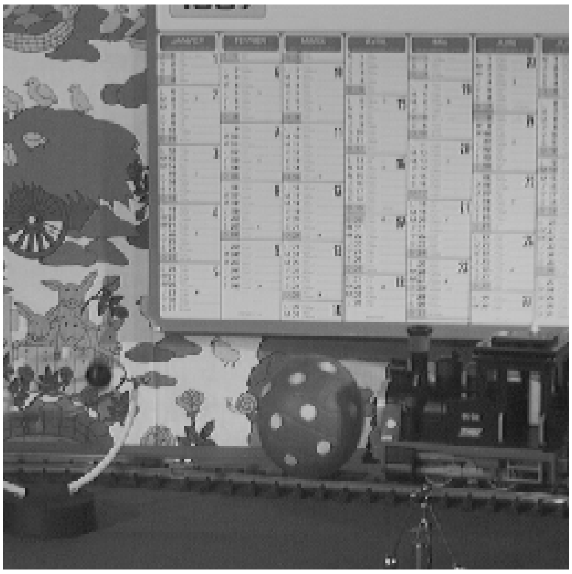

Figure 2. “Mobile & Calendar” sequence, converted to 256×256 format.

bit rates are embedded at the beginning of the bit-stream needed for the lossless coding.

5.3 Coding Results

The coding results are obtained for eight frames of the gray-scale “Mobile & Calendar” sequence, which we have converted to the 256×256 format. This is an ISO class C sequence, meaning high spatial detail and medium amount of movement. Fig. 2 depicts the first frame. To situate the coding performance of our wavelet video encoder, we compare it with the MPEG-4 Verification Model (*VM*) [3] which we put in unadvanced motion estimation mode. In this mode the encoder performs motion estimation with half pixel accuracy and uses 16×16 blocks. Since we have not implemented B-frames in our wavelet video encoder, the frame interdependency is restricted to IPPPPPPP. Table 2 contains the coding results for each frame. Since the “Mobile & Calendar” sequence is rectangular, no shape coding is required.

Frame no.	Image Type	PSNR (dB)	bpp (texture)
0	I	31.90	0.8597
1	P	30.57	0.2974
2	P	30.57	0.2890
3	P	30.57	0.2959
4	P	30.50	0.3028
5	P	30.57	0.2955
6	P	30.50	0.3063
7	P	30.50	0.2959

Table 2. MPEG-4 VM coding results for the “Mobile & Calendar” sequence (256×256).

While coding the sequence with our wavelet video encoder, we impose an identical number of bits per pixel (*bpp*) for each frame as for *VM*. This allows us to compare the reconstructed quality, expressed by PSNR values, to the quality of *VM*.

We compare *FS-AS/AD* to *FS-AD* for different block sizes, denoted by e.g. (2,4,8) representing 2×2 wavelet coefficients on decomposition level 3, 4×4 on level 2 and 8×8 on level 1. We use identical search ranges for these motion estimation algorithms, i.e. [-2,2] on level 3, [-4,4] on level 2 and [-8,8] on level 1. Experiments show that for the “Mobile & Calendar” sequence *AS*

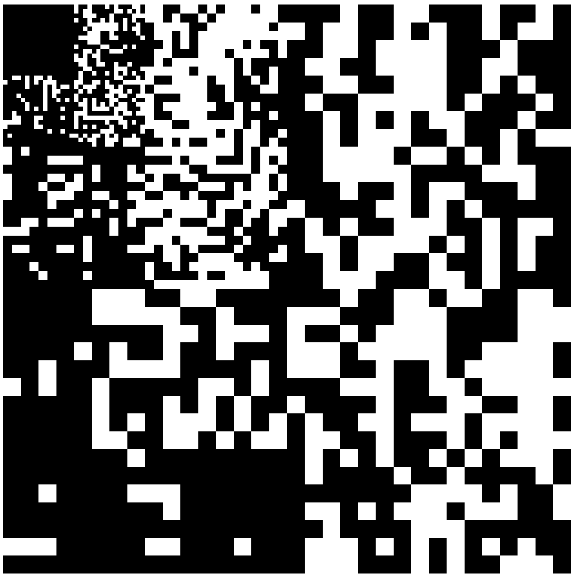

Figure 3. The *FS-AS/AD (2,4,8)* method attains a minimum with the *AS* criterion (white blocks) or *AD* criterion (black blocks).

reaches a smaller minimum than *AD* for more than half of the total number of blocks. This is illustrated in Fig. 3 which shows all blocks in the wavelet detail images. A block is drawn in white if the *AS* criterion reaches the lowest minimum or in black if the *AD* criterion attains the lowest minimal value.

To assess the coding gain obtained by performing motion estimation and compensation, we also coded the sequence on a frame by frame basis, i.e. complete intra frame coding of the sequence using the wavelet transform. The results are illustrated in Fig. 4.

Inter wavelet coding by using *FS-AD* or *FS-AS/AD* compared to intra wavelet coding, yields a considerable quality gain for the same number of bits per frame. The average gain attained by the worst *FS-AD* method, i.e. *FS-AD (8,8,8)*, is 1.7 dB. For the *FS-AD (4,4,4)* method, this is the best one, the average gain is 2.9 dB. If we compare the quality gains of the *FS-AS/AD* methods to intra wavelet coding, then we calculate an average gain of almost 2 dB for the worst method, i.e. *FS-AS/AD (8,8,8)*, and 3.4 dB for *FS-AS/AD (4,4,4)* which is the best one. Hence, we conclude that our wavelet video encoder achieves a considerable quality gain by performing motion estimation in the wavelet domain, compared to intra wavelet transform coding. Moreover, performing motion estimation in the wavelet detail images by using both the absolute sum and the absolute difference as block matching criteria in the *FS-AS/AD* method, results in a quality gain that varies between 0.3 and 0.5 dB compared to the *FS-AD* method, which only uses the absolute difference. In this way the *FS-AS/AD (4,4,4)* method gets close to the quality curve of the *VM*, but does not surpass it. This is due to the restriction that we impose the same number of bits for every frame as for *VM*. By using our own bit allocation we are able to exceed the *VM* curve. This is shown in Fig. 4 by the “*FS-AS/AD (4,4,4)* bit allocation” curve. We see that this curve is slightly above the *VM* curve for the inter wavelet coded frames. Moreover, the intra wavelet coded image is approximately 2 dB above the intra coded DCT image. Although we changed the bit allocation for this sequence, the total number of bits is still the same as for *VM*. This indicates that our wavelet video encoder needs its own bit allocation procedure to attain an optimal rate distortion result.

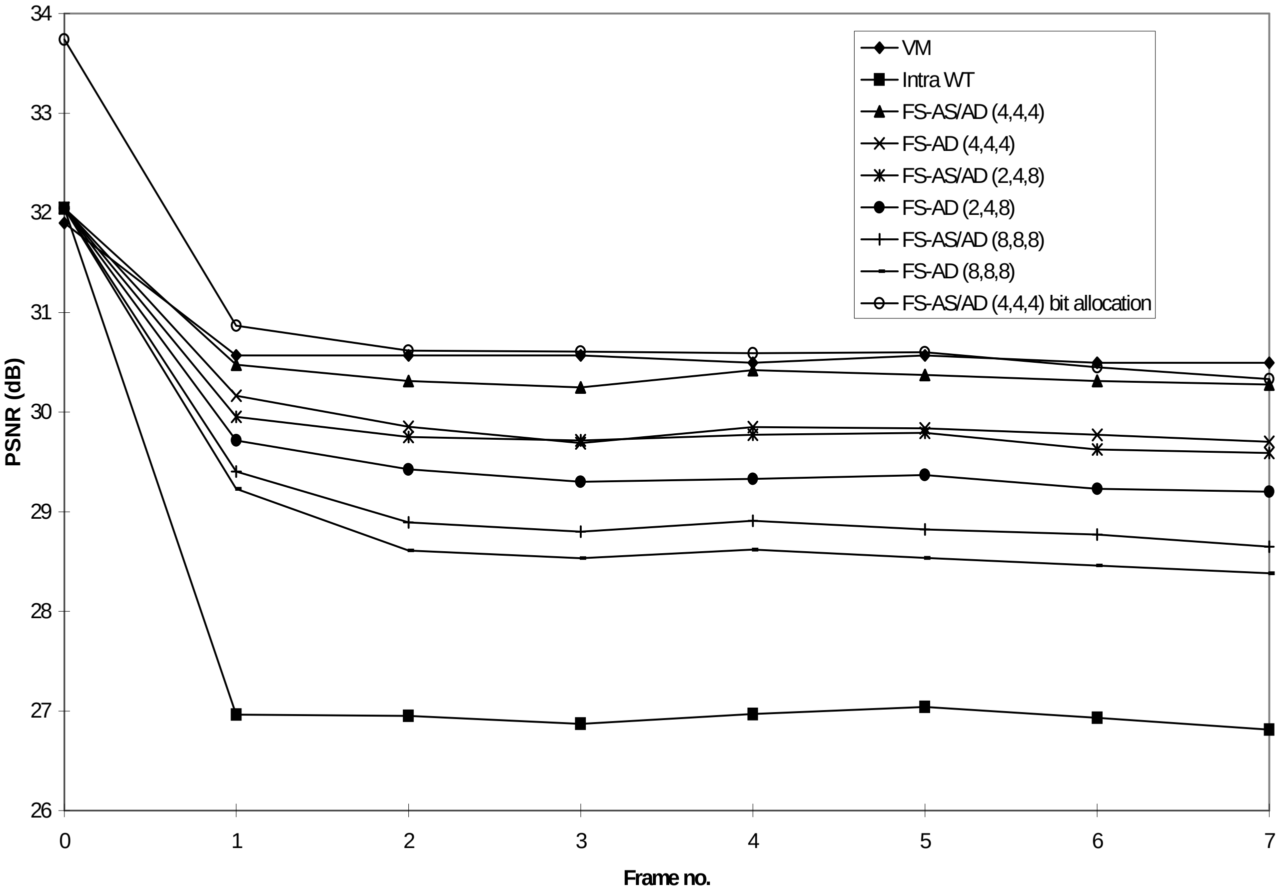

Figure 4. Coding results of the wavelet video encoder, using FS-AS/AD or FS-AD, and the MPEG-4 Verification Model (VM) for the "Mobile & Calendar" sequence. Intra wavelet coding of all frames is also indicated.

6 CONCLUSION

In this paper we have shown that motion estimation in the wavelet detail images using the combination of the absolute sum and the absolute difference, results in a significant image quality gain of up to 0.5 dB compared to using only the absolute difference for the same bit rate. By allocating bits to each frame in a more optimal way, our wavelet video encoder is able to slightly surpass the quality curve of *VM* for the inter coded images, and moreover a 2 dB gain is obtained for the intra coded image. We conclude that our wavelet video encoder, implementing *FS-AS/AD*, can be made competitive to current video coding techniques, while providing new features such as a simplified encoder structure and progressive transmission for variable channel bandwidths.

7 ACKNOWLEDGEMENTS

This research was funded by IWT: IT TeleVision Project (ITA/950202).

8 REFERENCES

[1] Cafforio C., Guaragnella C., Picco R. *Motion Compensation and Multiresolution Coding.* Elsevier Signal Proc. : Image Communication, Vol. 6, 1994, pages 123-142.

[2] Dufaux F., Moccagatta I., Kunt M. *Motion-Compensated Generic Coding of Video Based on a Multiresolution Data Structure.* Optical Engineering, Vol. 32, No. 7, July 1993, pages 1559-1570.

[3] ISO/IEC JTC1/SC29/WG11 N1902. *Information Technology - Coding of Audio-Visual Objects: Visual.* International Organization for Standardization (ISO), Fribourg, October 1997.

[4] Mandal M.K., Chan E., Panchanathan S. *Multiresolution Motion Estimation Techniques for Video Compression* (preprint).

[5] Martucci S., Sodagar I., Zhang Y.-Q. *A Zerotree Wavelet Video Coder.* IEEE Trans. on Circ. and Syst. for Video Techn., Vol. 7, No. 1, February 1997, pages 109-118.

[6] Munteanu A., Cornelis J., Van der Auwera G., Cristea P. *A Wavelet Based Lossless Compression Scheme with Progressive Transmission Capability.* Submitted for publication.

[7] Van der Auwera G., Lafruit G., Cornelis J. *Arithmetic Complexity of Motion Estimation Algorithms.* Proceedings of the IEEE Benelux Signal Processing Symposium, Leuven, Belgium, March 1998, pages 203-206.

[8] Van der Auwera G., Munteanu A., Lafruit G., Cornelis J. *Video Coding Based on Motion Estimation in the Wavelet Detail Images.* Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Seattle, May 1998.

[9] Zhang Y.-Q., Zafar S. *Motion-Compensated Wavelet Transform Coding for Color Video Compression.* IEEE Trans. on Circ. and Syst. for Video Techn., Vol. 2, No. 3, September 1992, pages 285-296.