VIDEO CODING BASED ON MOTION ESTIMATION IN 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 - VSDM, 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 uses two matching criteria, namely the absolute difference and the absolute sum. For a wavelet decomposed one-dimensional step function, it is shown that for odd translations of the step, the absolute sum reaches a smaller minimum than the absolute difference. We also derive in this case a constraint on the highpass filter coefficients so that a zero prediction error can be reached by using the absolute sum. Although this cannot be easily generalized for an arbitrary signal profile, experimental results obtained with photorealistic image sequences indicate that the prediction error can be reduced with respect to techniques that only use the absolute difference as matching criterion.

1. INTRODUCTION

In recent years, wavelets have proven to be successful in compressing still images. Compared to the classical DCT approach (JPEG), the wavelet based compression schemes have the advantage of a much better image quality obtained at very high compression ratios. Still image compression via the wavelet transform leads to a graceful image degradation at increased compression ratios, and does not suffer from the annoying block artefacts, which are typical for JPEG at very low bit rates. Another advantage of wavelets over DCT is the inherent multiresolution nature of the transformation, so that progressive transmission of images comes in a natural way. These advantages can be efficiently exploited for video sequences, especially in very low bit rate applications that can benefit from the improved image quality. Moreover, the progressive transmission capability is important to support variable channel bandwidths.

In this paper, we start by briefly discussing the use of wavelets in video compression. We indicate the problems associated to using the wavelet transform in a classical video codec as a substitute for the DCT. We describe a video encoder that does not contain an inverse wavelet transform. For this architecture we propose a new block based motion estimation and compensation technique which is able to reduce the prediction error with respect to the classical techniques. For an original signal consisting of a simple step function in the spatial domain, we indicate that in the wavelet domain it is better in certain cases to sum the predicted samples and the current samples to get the error samples, than to subtract them.

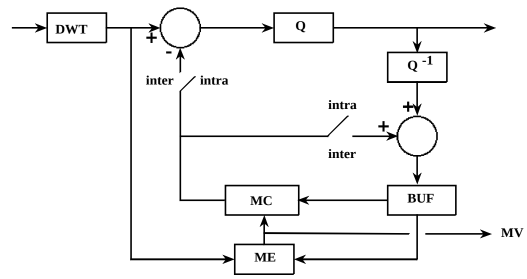


Figure 1. Wavelet based video encoder without inverse discrete wavelet transform.

Although this statement cannot be easily proven for an arbitrary signal profile, our experiments with image sequences show that a reduction of the prediction error is feasible. Therefore, we propose an algorithm that performs motion estimation in the detail images of the wavelet transform, by using two matching criteria, namely the absolute difference and the absolute sum.

The structure of this paper is as follows: in the next section we discuss the use of the wavelet transform in a video codec. Section 3 verifies our statement for a simple 1D step function, under the assumption of a particular type of wavelet filters. In sections 4 and 5 we respectively describe our algorithm and the results obtained with an image sequence.

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][4]. A drawback of this implementation is that for inter frame coding the wavelet transform is applied to the complete error image, which contains all the block artefacts. These artificial discontinuities, introduced in the motion field, lead to undesirable high-frequency subband coefficients that reduce the compression efficiency [2].

To avoid this limitation, the discrete wavelet transform is taken out of the prediction loop which results in the video encoder architecture [5] depicted in Fig. 1. Both motion estimation and compensation are performed in the wavelet domain, i.e. in the average image of the highest 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. The advantages of such a codec are: (1) the motion field blocking artefacts are no longer transformed to the wavelet domain and (2) no inverse discrete wavelet transform is needed, so that from a hardware point of view the encoder can be simplified.

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. Motion estimation and compensation are not as simple as in the spatial domain, where blocks are taken out of the reference image and are used to predict the next image. In the wavelet domain the required blocks are not directly available, therefore one cannot use the same techniques as in the spatial domain. 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 $k \cdot 2^J$, $k \in N$, where J denotes the number of decomposition levels. In this case, the same motion estimation and compensation approaches can be used in the wavelet domain as in the spatial domain.

Some methods have already been introduced in [3][5]. 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 matching criterion. To get the error image, the predicted wavelet image is subtracted from the original wavelet image (see Fig. 1), just as one would do in the spatial domain. 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.

We will show in the following sections that the prediction error of the detail images can be reduced if one considers both summing and subtracting the original and the predicted blocks.

3. 1D-STEP SHIFT COMPENSATION

The detail images contain high frequency information which corresponds mainly to edges in the spatial domain. To facilitate the calculations, we restrict the analysis to the one-dimensional case, and we model an arbitrary edge by a step profile.

Denote by h and g the filters used to perform a one-dimensional biorthogonal wavelet analysis of the step function $x(n)$. The lowpass filter h is symmetric around $n=0$, while the highpass filter g is symmetric around $n=-1$. Consider that $g(n)$ has $2N+1$ coefficients, and introduce the notation $\bar{g}(n)=g(n-1)$. The highpass component obtained from a one level wavelet analysis of $x(n)$ is given by:

$$x_g(n)=\bar{g}(0)x(2n+1)+\sum_{p=1}^{N}\bar{g}(p)\left[x(2n+1-p)+x(2n+1+p)\right].$$

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 $y_g(n)$.

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

Filters	AS	AD
Biorthogonal (2.4)	0.0000	0.3535
Biorthogonal (2.8)	0.0000	0.3535
Biorthogonal (3.9)	0.1768	0.3535
Biorthogonal (5.5)	0.1772	0.5459
Biorthogonal (6.8)	0.1315	0.4342
Biorthogonal (9.7)	0.0883	0.4349

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, if k is odd, the absolute sum between the predicted samples $x_g(n-s)$ and the original samples $y_g(n)$ is lower than the absolute difference, for specific values of s. We show this in the following for the particular case $s=1$.

It is easy to prove that the general case of an odd shift k can be restricted to the particular case $k=1$. We will assume $k=1$ in the remainder of this section, which leads to the highpass component $y_g(n)$ given by:

$$y_g(n)=\bar{g}(0)\cdot x(2n)+\sum_{p=1}^{N}\bar{g}(p)\cdot\left[x(2n-p)+x(2n+p)\right].$$

We denote by AD and AS the absolute difference respectively the absolute sum between the shifted wavelet component $x_g(n-s)$ and $y_g(n)$; the expressions of AD and AS for $s=1$ are:

$$AD=\sum_n\left|x_g(n-1)-y_g(n)\right|=\sum_n d(n),$$

$$AS=\sum_n\left|x_g(n-1)+y_g(n)\right|=\sum_n s(n).$$

Taking into account that $\bar{g}(n)=\bar{g}(-n)$, we derive:

$$d(n)=\left|\sum_{p=1}^{N+1}\left(\bar{g}(p-1)-\bar{g}(p)\right)\cdot\left(x(2n-1+p)-x(2n-p)\right)\right|,$$

$$s(n)=\left|\sum_{p=1}^{N+1}\left(\bar{g}(p-1)+\bar{g}(p)\right)\cdot\left(x(2n-1+p)+x(2n-p)\right)\right|,$$

where $\bar{g}(N+1)=0$. Since the input signal $x(n)$ is a step function, it can be proven that $d(n)=d(-n)=\left|\bar{g}(2n)\right|$, for every value n, verifying $0\le n\le\left[N/2\right]$.

The expression for $s(n)$ is:

$$s(n)=s(-n)=\left|\bar{g}(2n)+2\cdot\sum_{p=1+2n}^{N}\bar{g}(p)\right|,\ \forall n\in N, 0\le n\le\left[N/2\right].$$

If $n<-\left[N/2\right]$ or $n>\left[N/2\right]$, we can show that $s(n)=d(n)=0$.

Finally, the absolute difference and the absolute sum are given by:

$$AD = \left|g(-1)\right| + 2\sum_{n=1}^{[N/2]}\left|g(2n)\right| \text{ , } AS = 2\sum_{n=1}^{[N/2]}\left|g(2n-1) + 2\sum_{p=2n}^{N-1} g(p)\right| .$$

The values of *AD* and *AS* are evaluated for different biorthogonal filter banks. As we note from table 1, 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 if the filter coefficients satisfy the constraint:

$$\left|g(-1)\right| + 2\sum_{n=1}^{[N/2]}\left|g(2n)\right| = 0 .$$

Similar calculations are made to derive *AD* and *AS* for $s \neq 1$. For all the tested filters, the minima of the absolute sum are reached in $s = 1$, and they are smaller than the minima of the absolute difference. An example is given in Fig. 2, that depicts *AD* and *AS* as a function of *s*, for the biorthogonal filters (2.4) and (5.5). The same conclusion can be formulated from Fig. 3 in the case of the biorthogonal filter (9.7). It results that the smallest prediction error is attained if $y_g(n)$ is predicted from $x_g(n-1)$ by using the *AS* criterion.

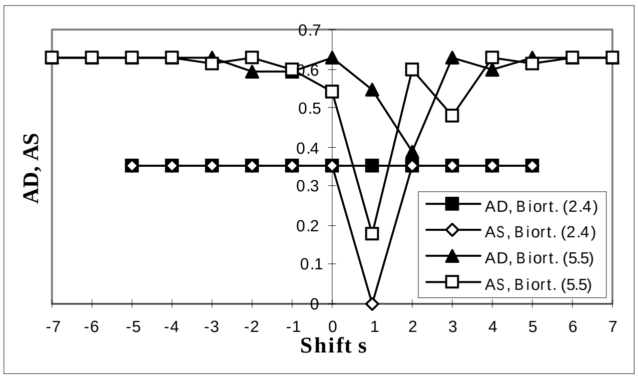


Figure 2. *AD* and *AS* as a function of *s*, for the biorthogonal filters (2.4) and (5.5).

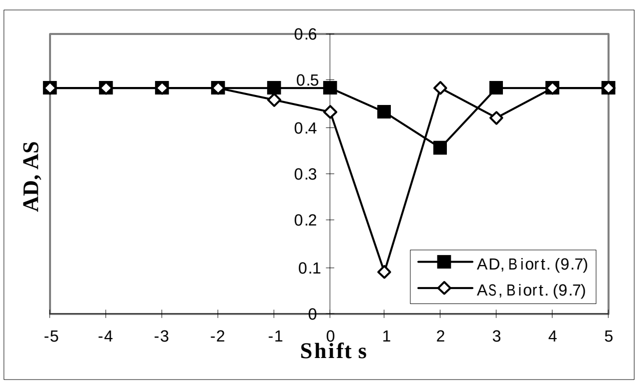


Figure 3. *AD* and *AS* as a function of *s*, for the biorthogonal filter (9.7).

4. ALGORITHM DESCRIPTION

In the above section we have shown that for odd shifts of the step function a small or even zero prediction error can be found if the predicted wavelet coefficients are summed to the original coefficients. If the shift is even, then we have to subtract them to get a zero error. Unfortunately, these simple rules are not completely valid for a general signal profile. However, experiments on photorealistic image sequences suggest that the prediction error can still be reduced by using these principles. In this section we describe an algorithm that performs motion estimation in the wavelet detail images by using two matching criteria, namely *AD* and *AS*. We will compare the resulting prediction error of our algorithm with the minimal error that can be reached by just using *AD* as a matching criterion.

4.1 FS using Absolute Difference

This method performs full-search motion estimation on every level of the wavelet decomposition by using *AD* as error criterion, and calculates the error image by subtracting the predicted wavelet image from the original image.

In our simulations, we use a 3-level 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 any detail image, while in the second one 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 the *FS-AD* method (full-search using *AD*) in the section reporting the experimental results.

4.2 FS using Absolute Sum and Difference

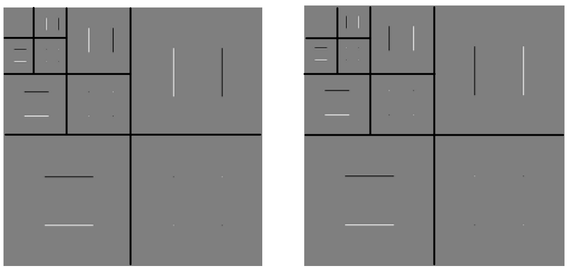
We propose a motion estimation algorithm that performs full-search motion estimation on every level of the wavelet decomposition and implements two matching criteria for finding the best block, namely *AS* and *AD*. The block sizes in every level and the search ranges are specified as in the *FS-AD*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

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. If this is the case, then the motion vector is determined by the position of the block in the reference image where *AS* is minimal. Conversely, if the minimum of *AD* is the smallest, then the motion vector will be the same as for the *FS-AD* method. 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*).

One bit for each predicted block has to be recorded as side information, to distinguish between the motion vectors determined by the *AD*, respectively by the *AS* criterion. 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 was retained, so that when the predicted wavelet image is subtracted from the original image these blocks are summed.

5. RESULTS

The theory from section 3 is illustrated with an artificial sequence of two images: a square, respectively the same square shifted to the right with one pixel. Fig. 4 shows the 3 levels wavelet decomposition performed with the (2,4) biorthogonal filters on (a) the original square, respectively (b) the shifted square. The shift variance is clearly visible, because the values of the vertical edges change due to the shift.



(a) (b)

Figure 4. The detail images obtained with a (2,4) biorthogonal wavelet analysis of (a) the square and (b) the square shifted to the right with one pixel.

If we apply the *FS-AD* method to this sequence, we find an error image which has a *MSE* of 10.07. For this experiment we used dyadic block sizes of $2^{4-j} \times 2^{4-j}$ coefficients, or in shorter notation *(2,4,8)* - corresponding to *j* values *(3,2,1)*. The maximum search ranges in the detail images are 2^{4-j} pixels. If we apply the *FS-AS/AD* method, then we find a *MSE* of 0.84, which implies that our method enables a nearly perfect prediction of the shifted square.

We use the well-known “Mobile Calendar” sequence to illustrate that our *FS-AS/AD* method also reduces the *MSE* of the error image in photorealistic sequences. The images out of the sequence are transformed with the (9,7) biorthogonal wavelet filters and are motion compensated to generate the error images. The motion estimation method uses the preceding image as a reference for the current image. Fig. 5 shows the *MSE*’s obtained with the methods *FS-AD* and *FS-AS/AD*, calculated for different block sizes.

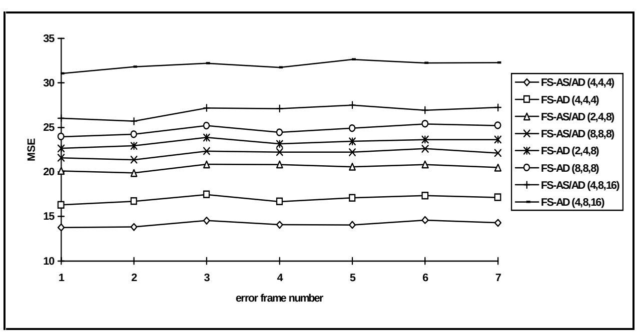


Figure 5. MSE results for the “Mobile Calendar” sequence.

It is clear from this figure that our *FS-AS/AD* method results in smaller MSE values than *FS-AD* for the same block sizes.

The experiments also show that for the “Mobile Calendar” sequence the *AS* reaches a smaller minimum than the *AD* for more than half of the total number of blocks. This is illustrated in Fig. 6 which shows all blocks on every level. A block is drawn in white if the *AS* criterion reaches the lowest minimum or in black if the *AD* criterion attains the lowest minimum.

Figure 6. The *FS-AS/AD (2,4,8)* method attains a minimum with the *AS* criterion (white blocks) or *AD* criterion (black blocks).

6. SUMMARY

In this paper we show that motion estimation in the wavelet detail images using the combination of the absolute sum and the absolute difference reduces the MSE of the error image. We indicate that for odd shifts of a one-dimensional step function the absolute sum leads to a zero prediction error if a specific constraint on the wavelet filter coefficients is respected. Although this constraint cannot be easily extended for a general signal profile, as our experiments point out, an additional reduction of the prediction error is obtained with respect to techniques which use only the absolute difference as a matching criterion.

7. REFERENCES

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

[2] Dufaux F., Moccagatta I. and Kunt M. “Motion-Compensated Generic Coding of Video Based on a Multiresolution Data Structure”. *Optical Engineering*, 32(7):1559-1570, 1993.

[3] Mandal M.K., Chan E. and Panchanathan S. “Multiresolution Motion Estimation Techniques for Video Compression” (preprint).

[4] Martucci S., Sodagar I. and Zhang Y.-Q. “A Zerotree Wavelet Video Coder”. *IEEE Trans. on Circ. and Syst. for Video Techn.*, 7(1):109-118, 1997.

[5] Zhang Y.-Q. and Zafar S. “Motion-Compensated Wavelet Transform Coding for Color Video Compression”. *IEEE Trans. on Circ. and Syst. Video Techn.*, 2(3):285-296, 1992.