AN OVERCOMPLETE DISCRETE WAVELET TRANSFORM FOR VIDEO COMPRESSION

Nicu Sebe, Claudiu Lamba, Michael S. Lew

Leiden Institute of Advanced Computer Science,

Niels Bohrweg 1, 2333 CA, Leiden, The Netherlands

ABSTRACT

The translated function with any integer multiple of the sampling period is completely represented in the wavelet space by one of the Overcomplete Discrete Wavelet Transform ($ODWT$) members. This theoretical result leads to a new motion estimation and motion compensation scheme working in the wavelet transform domain. Our experiments, performed on real image sequences, show high quality and low bit rate performances. Moreover, by performing the motion estimation in the wavelet space a major reduction of the computational cost is achieved.

1. INTRODUCTION

The multiresolution representation of videos has the advantage of scalability, i.e. the possibility to transmit the same sequence at different resolution as high resolution television, normal television, videophone, and videoconferencing. In addition, this representation seems to be the method of decomposition and understanding of images in the human visual system [1]. Among different possibilities of multiresolution analysis and synthesis, wavelet functions are the most adapted to these purposes due to their scaling and translation properties.

The present difficulty in obtaining the motion compensated image from the multiresolution representation is caused by the impossibility to obtain a right answer by inverting the operators of translation and wavelet transform. Our goal is to give a compression algorithm in the multiresolution wavelet space for motion compensation on an arbitrarily dense set of position-scale samples. To do this we should know the values of the multiresolution representation of the function (the transform coefficients) in some other points than those given by the transform. The problem may appear in practical cases when both the function and wavelets are discrete and the samples of the translated function do not generally correspond with the translated coefficients in the wavelet space. In particular, only when the translation is an integer multiple of the corresponding resolution scale, the wavelets are translated versions of the originals.

In video coding, several types of interframe predictions are used to remove the temporal interframe redundancy. Motion compensation has been used as an efficient scheme for temporal prediction. In order to perform motion compensation in the wavelet domain, block matching can be applied to the wavelet coefficients [2, 3, 4]. However, motion compensation in the wavelet domain is highly dependent on the alignment of the signal and the discrete grid chosen for the analysis. There exist very large differences between the wavelet coefficients of the original image and the one-pixel-shifted image. The shift-variant property happens frequently around the image edges, so motion compensation of the wavelet coefficients can be difficult.

To overcome the shift-variant property of the discrete wavelet transform, Kim and Park [5] use a low-band-shift (LBS) method. Cafforio, et al. [6] use a version of the translated image reconstruction in the wavelet space in the context of motion compensation. The proposed method is based on the possibility to obtain the wavelet functions of the translated original by autocorrelation and crosscorrelation of the scaling and wavelets functions. Other methods to obtain the property of signal reconstruction in arbitrary position and/or scales are the shiftable multiscale transforms [7], the reconstruction of the function from zero crossing [8], and "a trous" algorithm given in [9].

The methods previously described in the literature encounter some major drawbacks when they are used as compression methods based on a multiresolution image representation and wavelet transform. Namely, the shiftable multiscale transform [7] limits the analysis functions to those satisfying a particular matrix equation, and for real filters it is possible to find only an approximate solution. The zero crossing method [8] despite giving good signal reconstruction, is rather complicated, necessitating an iterative algorithm.

Based on the results presented by Rioul and Duhamel [10] we propose another method (similar to the Laplacian pyramids [14]) theoretically developed in this work. Practical implementation of the method in the context of image compression by motion estimation and motion compensation is also given. The reminder of the paper is organized as follows. Section 2 demonstrates the possibility to reconstruct an arbitrarily displaced image from one particular realization of the so called "The Overcomplete Wavelet Transform." Section 3 presents the results obtained on some well-known image sequences. Conclusions are given in Section 4.

2. THE MULTIRESOLUTION IMAGE DECOMPOSITION

The wavelet transform (WT) of a function $f(x) \in L_2(R)$ is defined [11] as

$$WT\{f(x); a, b\} = \int f(x)\Psi_{a,b}(x)dx \tag{1}$$

where the functions $\Psi_{a,b}(x)$ define the family of the wavelet functions, with $a \neq 0$ the scale of the transform and b the spatial location. The function $f(x)$ may be completely characterized by its wavelet coefficients $C_{j,k}$. These coefficients are the WT of the function $f(x)$ on a dyadic scale $a = 2^j$ and discrete translations $b = 2^j k$, and form a Discrete Scale Wavelet Transform ($DSWT$):

$$DSWT\{f(x); 2^j, 2^j k\}=\{C_{j,k}\}=WT\{f(x); a=2^j, b=2^j k\} \tag{2}$$

The wavelet function forms a dyadic family:

$$\Psi_{j,k}(x) = 2^{-\frac{1}{2}j}\Psi(2^{-j}x - k) \tag{3}$$

which is orthogonal and complete. Thus the function $f(x)$ may be obtained from its wavelet coefficients $C_{j,k}$:

$$f(x) = \sum_{j\in Z}\sum_{k\in Z} C_{j,k}\Psi_{j,k}(x) \tag{4}$$

For discrete signals $f(n), n \in Z$, the Discrete Wavelet Transform (DWT) is defined as:

$$DWT\{f(n); 2^j, 2^j k\} = \{c_{j,k}\} = \sum_{n\in Z} f(n)g_j^*(n - 2^j k) \tag{5}$$

with $g_j(n-2^jk)$ the discrete equivalent of the $\Psi_{j,k}(x)$.

It is easy to demonstrate that all continuous wavelets of the arbitrary translated function $f(x+x_0)$ are translated wavelets with the same quantity of the function $f(x)$. When both the function and wavelets are discrete, the samples of the translated function $f(x+x_0)$ do not correspond with the translated coefficients representing the DWT unless the translations $x_0 = k2^{j_0}$ are considered.

2.1. The Overcomplete Wavelet Transform

The problem we have to solve is similar to the so called "initialization" problem [12], that is the approximation of the $DSWT$ by computing the calculable DWT. Consider

$$\hat{f}(x) = \sum_n f(n) i(x-n) \approx f(x) \tag{6}$$

and the discrete filter $g_j(n)$

$$\sum_n g_j(n)\phi(x-n) = 2^{-\frac{j}{2}}\hat{\Psi}_j(x) \tag{7}$$

where $\hat{f}(x)$ and $\hat{\Psi}_j(x)$ are the approximated function and wavelet, respectively, and $i(x)$ and $\phi(x)$ are the corresponding interpolation functions.

The DWT is defined by the approximated coefficients:

$$\hat{C}_{j,k}=DWT\{f'(n);2^j,2^jk\}, \text{ with } f'(n)=\sum_m f(m)l(n-m) \tag{8}$$

The function $l(n)$ results by substitution of $\hat{f}(x)$ with $\hat{\Psi}_j(x)$ in the definition of the $DSWT$:

$$l(n) = \int i(x)\phi^*(x-n)dx \tag{9}$$

Thus the coefficients $\hat{C}_{j,k}$ are the DWT of a prefiltered version of $f(n)$ with a filter $l(n)$ given by (9) when $f(x) = \hat{f}(x)$ and $c_{j,k} = \hat{C}_{j,k}$.

Our interest is the possibility to approximate a nearly continuous WT representation in the time-scale plane for obtaining a good estimation of the motion. The algorithm may be extended to the required computation of the WT at scale 2^j but in every k points:

$$C_j^k = WT\{f(x);2^j,k\} = DWT\{f'(n);2^j,k\} \tag{10}$$

where, as before, $f'(n)$ is a prefiltered discrete input. The difference is that now the DWT is computed for all integer values of $b=k$ and not only in every $b=2^jk$. We call this representation an Overcomlete Discrete Scale Wavelet Transform ($ODSWT$):

$$\begin{aligned} ODSWT\{f(x);2^j,k\} &= C_j^k = WT\{f(x);2^j,k\} \\ &= 2^{-\frac{j}{2}}\int f(x)\Psi(2^{-j}(x-k))dx \end{aligned} \tag{11}$$

with denser wavelet functions:

$$\Psi_j^k = 2^{-\frac{j}{2}}\Psi(2^{-j}(x-k)),\ j,k \in Z \tag{12}$$

Similarly, for a discrete function $f(n)$, the Overcomplete Discrete Wavelet Transform ($ODWT$) is:

$$ODWT\{f'(n);2^j,k\} = \{c_j^k\} = \sum_{n\in Z} f(n)g_j^*(n-k) \tag{13}$$

The Eq. (11) shows the redundancy of the $ODSWT$. It can be demonstrated that $ODSWT\{f(x);2^j,k\}, j,k \in Z$ is equivalent with the set $\{DSWT\{f(x);2^j,2^jk+m\}\}, m = 0,\ldots,2^j-1$ and $j,k \in Z$.

In this way, for every m we have a complete representation of the function f or, in other words, from every $DSWT\{f(x);2^j,2^jk+m\}$ with a given value m, we may obtain the function $f(x)$. It is easy to observe that for different values of m the translated versions of the dyadic scale WT are obtained.

Similarly, for the discrete equivalent of the function $f(x)$, a member of the Overcomplete Discrete Wavelet Transform can be written as:

$$\{c_j^{k,m}\} = DWT\{f'(n);2^j,2^jk+m\}, m = 0,\ldots,2^j-1 \tag{14}$$

If the function $f'(n)$ has the discrete wavelet transform:

$$DWT\{f'(n);2^j,2^jk\} = \{c_{j,k}\} \equiv \{c_j^{k,0}\} \tag{15}$$

the translated function $f'(n+m_0)$, with any integer m_0, has the DWT one of the $ODWT$ members:

$$\begin{aligned} DWT\{f'(n+m_0);2^j,2^jk\} &= \sum_{n\in Z} f'(n+m_0)g_j^*(n-2^jk) \\ &= ODWT_{m=m_0}\{f'(n);2^j,k+2^{-j}m_0\} = \{c_j^{k,m_0}\} \end{aligned} \tag{16}$$

The relation (16) demonstrated that: **The translated function with any integer multiple of the sampling period is completely represented in the wavelet space by one of the $ODWT$ members.**

Corollary: **It is always possible to obtain any translated version of the original discrete function from one of the $ODWT$ members.**

2.2. The coding scheme

Our coding method is based on the theoretical result we derived in the previous section. A frame $\mathbf{I}$ in an image sequence is represented in the multiresolution space by the entire set of the $ODWT$ members, $\{DSWT\}_m, m = 0,\ldots,2^j-1$. Using a full search block-matching procedure between each member of the set $\{DSW\mathbf{I}\}_m$ and the previous transmitted image $DSW\mathbf{I}_{-1}$, the member of the $ODWT$ that gives the minimum absolute difference is chosen. Let $m = v$ be this choice. All v in different subbands and resolution scales form the vector $\mathbf{v}$. This $\mathbf{v}$ selects the proper $ODWT$ member for coding. It is also necessary to obtain the value k representing the coarse displacement, in this case an integer multiple of 2^j. The values $\mathbf{v}$ and k are then lossless coded; the vector compression ratio attainable by these means is the order of 1.7.

An additional difference between $DSW\mathbf{I}_{-1}$ and $\{DSW\mathbf{I}\}_{m=v}$ is transmitted. A vector quantization is applied to this difference image. At the decoder, the wavelet coefficients translated by k and the codebook vector are added. The value of $\mathbf{v}$ is used to select the proper interpolation in DWT^{-1}.

3. RESULTS

In order to verify the proposed method, comparison studies were performed for the proposed method and two other methods: the conventional Full-Search Block-Matching Algorithm (FS-BMA) and the method proposed by Kim and Park [13] based on low-band shift (LBS).

The FS-BMA partitions each frame into a number of equal-sized blocks and finds a constant motion vector for each block by searching for its peak correlation with an associated block in the previous frame. In the LBS method [13] for motion estimation and compensation in the wavelet domain, the reference frame is shifted by one pixel along the x, y, and the diagonal directions, respectively, in the spatial domain. The shifted frames are transformed in the wavelet domain for motion estimation. These shift and wavelet-transform processes form the low-band-shift method. The next-level low-band-shift operations are repeated iteratively to the low-low band of each level. This low-band-shift method avoids the shift-variant property of the wavelet transform.

For our method, the input image is first decomposed into a pyramid structure with three levels using the $ODWT$ (Figure 1). The size of the blocks that were used for motion estimation purposes is $N \times N$, where $N = 2^{4-j}$ and $j = 1, 2, 3$ for each horizontal, vertical, and diagonal orientation band at the j^{th} level. The maximum search displacement is assumed to be 5 on the 1^{st} level, 3 on the 2^{nd}, and 2 on the 3^{rd} level and in the approximation subband. For the FS-BMA simulation, 8×8 subblocks are employed. The search window covers a 8 pixels maximum displacement.

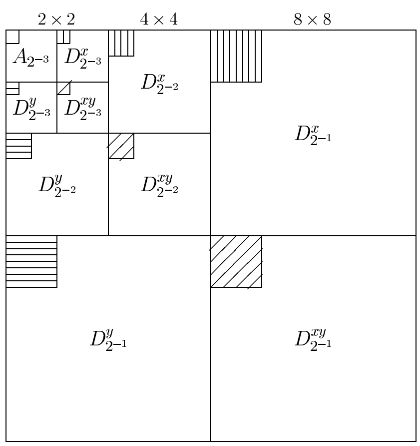


Figure 1: Composition of the block matrix

Coding simulations were performed on two well-known motion sequences: ”Flower Garden” and ”Football”. As an objective measure of the reconstructed image quality the peak-to-peak signal-to-noise ratio (PSNR) was considered.

Coding experiments for a succession of 7 frames at a bit rate of 0.14 bpp yield an average PSNR of 23.39 dB, 31.89 dB, and 28.03 dB for the FS-BMA, LBS, and our method ($ODWT$), respectively. A spectacular improvement of our algorithm is achieved by transmitting an additional 0.1-0.15 bpp VQ coded difference image. The simulation results for a succession of 7 coded frames are presented in Table 1. In this case, the penalties that has to be paid for the improvement in the image quality consist of an increase of the bit rate and extra computational complexity. Note that the results obtained with our method are better than the ones obtained with the LBS method. Moreover, the advantage of our method relies in its simplicity. We avoid any extra manipulation of the image in the spatial and wavelet domains required by the LBS method.

Rate distortion curves obtained by using the proposed scheme and the LBS method are displayed in Figure 2. This figure illustrates the increase of the PSNR as a function of the overall bit rate for a coded frames with respect to the additional transmission of the VQ-coded difference image. The average improvement of our method over the LBS method is 1.3 dB in PSNR.

Frame No.	Flower Garden PSNR (dB)		Football PSNR (dB)	
	$ODWT$	LBS	$ODWT$	LBS
1	38.23	36.14	35.33	35.13
2	37.01	35.89	35.12	35.03
3	35.65	35.52	34.83	34.53
4	34.89	34.71	34.26	33.81
5	34.50	34.12	33.85	32.97
6	33.81	33.51	32.93	32.65
7	32.86	33.09	32.43	31.83

Table 1: The values of the PSNR for 7 adjacent frames

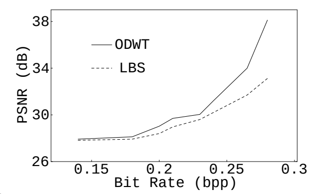


Figure 2: PSNR as a function of the overall bit rate

In Figure 3 two frames from the original sequence ”Flower Garden” with the corresponding reconstructed frames and the error images for each of them are presented. The reconstructed frames from the proposed method have good image quality and do not have blocking effects.

The computational requirement for a 8×8 search block of the FS-BMA is as follows. For a single search point, 64 subtractions, 64 absolute operations, and 63 additions are required. For the proposed method, the following operations are performed for a single search point. At level 1, 64 subtractions, 64 absolute operations, and 63 additions, at level 2, 16 subtractions, 16 absolute operations, and 15 additions, at level 3, 4 subtractions, 4 absolute operations, and 3 additions for each subblock are needed. In the approximation subband 4×64 subtractions, 4×64 absolute operations, and 3×64 additions are required.

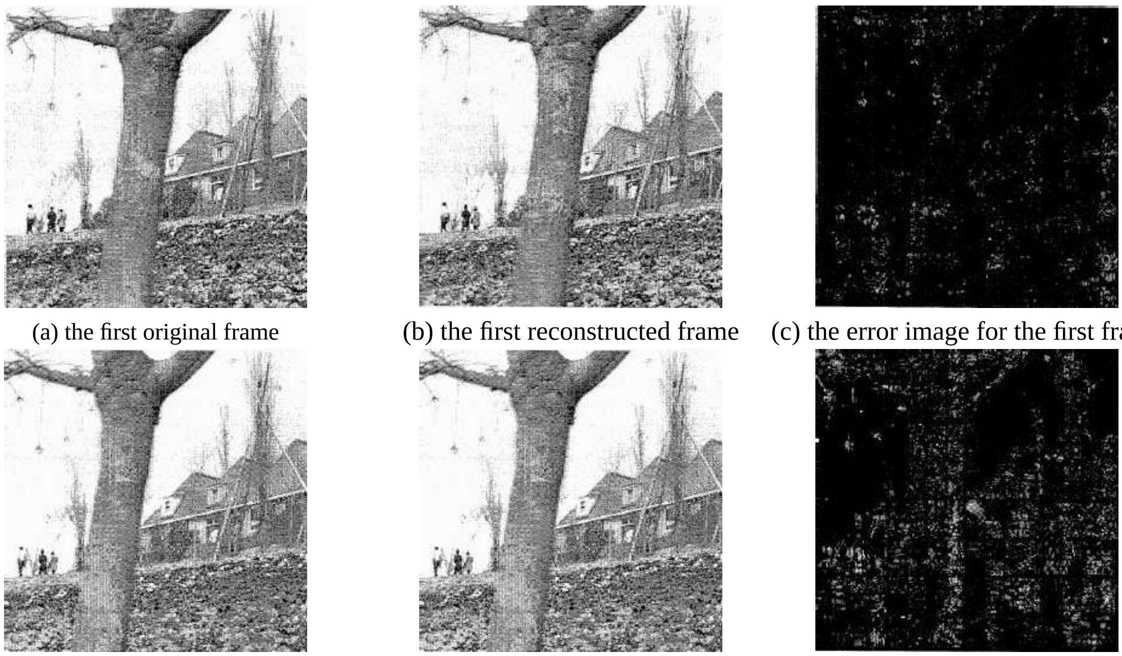
In Table 2 is presented the total number of operations performed for the entire image of 256×256 pixels in both cases. The comparison between FS-BMA and our method is made with respect to the number of searched blocks and the maximum displacement considered for each subblock. It can be seen that the computational burden for the proposed algorithm is about 42% of that of the FS-BMA method. Note that the number of operations performed for the LBS method is practically the same with the number of operations required for our method.

Method	Subtractions	Abs. operations	Additions
FS-BMA	18,939,904	18,939,904	18,643,968
$ODWT$	8,264,704	8,264,704	7,705,344

Table 2: Comparison of total number of operations

4. CONCLUSIONS

In this paper a new concept called ”The Overcomplete Discrete Wavelet Transform” was presented. We demonstrated the possibility to obtain for a given function f in the wavelet space, the



(a) the first original frame (b) the first reconstructed frame (c) the error image for the first frame

(d) the 7^{th} original frame (e) the 7^{th} reconstructed frame (f) the error image for the 7^{th} frame

Figure 3: An example of two coded frames using the $ODWT$ method

displaced function of f with any integer value of the sampling period. Based on this theoretical result we elaborated a compression method for image sequences. The motion estimation and motion compensation are both accomplished in the wavelet domain.

The experimental results on two well-known video sequences prove the possibility to transmit good quality images (PSNR 31.83-38.23 dB) at low bit rates (0.14-0.28 bpp). The results show that the proposed method outperforms Kim and Park's low-band-shift algorithm. Our method provides not only better results but it is also much simpler. We avoid any extra manipulation of the image in the spatial and wavelet domains required by the LBS method.

We consider that the proposed method gives a better matching to both properties of the image and the human visual characteristics as well. It also provides a better distribution of the information contained in the motion vector in low and high resolution images with advantages in transmission of image sequences at different bit rates.

5. REFERENCES

[1] J. Daugman, *Two-dimensional spectral analysis of cortical receptive profile*, Vision Research, 20:847-856, 1980.

[2] M. K. Mandal and S. Panchanathan, *Integrated compression and indexing of video in the wavelet domain*, Journal of Electronic Imaging, 9(2):85-100, 2000.

[3] K. Shen and E. Delp, *Wavelet-based rate scalable video compression*, IEEE Trans. Circ. Sys. Video Techn., 9(1):109-122, 1999.

[4] Y.-Q. Zhang and S. Zafar, *Motion-compensated wavelet transform coding for color video compression*, IEEE Trans. .Circ. Sys. Video Techn., 2(3):285-296, 1992.

[5] H.S. Kim and K.T. Park, *Wavelet-based moving-picture coding using shift-invariant motion estimation in wavelet domain*, Sign. Process.: Imag. Communic., 16(7):669-679, 2001.

[6] C. Cafforio, C. Guaragnella, F. Bellifemine, A. Chimienti, and R. Picco, *Motion compensation and multiresolution coding*, Sign. Process.: Imag. Communic., 6(2):123-142, 1994.

[7] E.P. Simoncelli, W.T. Freeman, E.H. Adelson, and D. J. Hegger, *Shiftable multiscale transforms*, IEEE Trans. Inf. Theory, 38(2):587-607, 1992.

[8] S. Mallat, *Zero-crossing of a wavelets transform*, IEEE Trans. Inf. Theory, 37(4):1019-1033, 1991.

[9] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, *A real-time algorithm for signal analysis with the help of the wavelet transform*, Wavelets, Time-frequency Representations, and Phase Space, Springer Verlag, 1989.

[10] O. Rioul and P. Duhamel, *Fast algorithms for discrete and continuous wavelet transforms*, IEEE Trans. Inf. Theory, 38(2- Part II):569-586, 1992.

[11] C.K. Chui, An introduction to wavelets, Academic Press, 1992.

[12] M.J. Shensa, *The Discrete Wavelet Transform: Wedding the 'A-Trous' and Mallat algorithms"*, IEEE Trans. Sign. Process., 40(10):2464-2482, 1992.

[13] K.T. Park and H.S. Kim, *Motion estimation using low-band-shift method for wavelet-based moving-picture coding*, IEEE Trans. Imag. Process., 9(4):577-587, 2000.

[14] P.J. Burt and E.H. Adelson, *The Laplacian pyramid as a compact image code*, IEEE Trans. Communic., 31:532-540, 1983.