

# On the Performances and Improvements of Motion-Tuned Wavelets for Motion Estimation.

BRAULT P.

LSS, Laboratoire des Signaux et Systemes, CNRS UMR 8506,  
Supelec, Plateau du Moulon, 91192 Gif sur Yvette Cedex, FRANCE.

IEF, Institut d'Électronique Fondamentale, CNRS UMR 8622.  
Université Paris-Sud, Bâtiment 220, 91405 Orsay Cedex, FRANCE.

*Abstract* : - The Spatio-Temporal Motion-Tuned Wavelet Transform is a construction started around the early 90's and that has been investigated in particular for motion-tracking purposes. Recently this approach has been reviewed in new applications of motion estimation applied either to video compression or to scene analysis. We show in this paper some performances reached with such constructions and make a comparison with another close construction based on the optical flow computation. We also discuss improvements of the construction of the MTSTWT and show which results can be reached with this interesting wavelets family.

*Key-Words*:- Motion-Tuned Spatio-Temporal Wavelet Transform, Continuous W.T., Motion Detection, Motion Estimation, Video Compression, Scene Analysis, Object Tracking, Optical Flow, Fast algorithm, Matching Pursuit, Best Basis.

## 1 Introduction

Wavelets are today a very "hot" subject in video compression. Several developments on the basis of the MPEG4 standard have been proposed. The result of a recent "Call for Evidence" in July 2003, Trondheim, has shown a good behaviour of compression schemes based on wavelets. The "Call for proposal" at the end of 2003 will probably show the emergence of wavelets within the upcoming standards. A group using wavelet for compression has been constituted within the MPEG4 team.

In the approach we have been investigating recently for motion detection and estimation, as well as for video compression and scene analysis, objects motions are tracked and quantified by wavelets tuned to motion [3]. We have described a scheme where objects trajectories can be computed on the basis of the motion coefficients provided by the wavelet transform, the so-called MTSTWT, and an a priori model for the trajectory (Nth order polynomial, Spline etc.). So motion estimation and temporal redundancy reduction (especially in compression applications) could be based on

the computed trajectories.

This approach differs from other recent use of wavelets for new video compression proposals like :

- 1) Hybrid compression with "post motion-compensation filtering" with wavelets.
- 2) 2D+T compression with motion compensation and lifting [7].
- 3) Fast optical flow computation resolved on a 2D orthogonal basis [1].

This is not an exhaustive description of the recent application of wavelets in video compression and motion estimation, but we will concentrate here on the "unwarpped" signal analysis with a specific class of motion-tuned wavelets that we have started to investigate [2, 3].

## 2 Construction of Spatio-temporal Motion-Tuned Wavelets

We recall here the basis of the construction of spatio-temporal motion-tuned wavelets described in [5, 8, 10].

The definition of a wavelet transform tuned to motion is

done by means of the **composite operator**  $\Omega_g$ . This one is defined by the application of transform operators on a mother wavelet:

$$[\Omega_g\psi](\vec{x}, t) = [T^{\vec{b}, \tau} R^\theta \Lambda^c D^a \psi](\vec{x}, t) \quad (1)$$

where the transform operators are respectively the spatio-temporal translation, the rotation, the velocity tuning and the scaling.

This gives, when replacing the operators by their expressions, in the Euclidian or spatio-temporal domain (direct space):

$$[\Omega_g\psi](\vec{x}, t) = a^{-3/2} \psi \left( \frac{c^{-1/3}}{a} r^{-\theta} (\vec{x} - \vec{b}), \frac{c^{2/3}}{a} (t - \tau) \right) \quad (2)$$

We can use the above relation with a spatio-temporal morlet wavelet defined by :

$$\psi(\vec{x}, t) = (e^{-1/2|\vec{x}|^2} \cdot e^{ik_0\vec{x}}) \times (e^{-1/2t^2} \cdot e^{i\omega_0 t}) \quad (3)$$

This last expression is the simplified expression of the admissible Morlet wavelet, under the assumption that  $k_0 \geq \pi\sqrt{\frac{2}{Ln2}} \simeq 5.336$  and  $\omega_0 \geq \pi\sqrt{\frac{2}{Ln2}} \simeq 5.336$ , which enables to neglect the admissibility terms [5, 6].

The previous expression enables to define a wavelet transform tuned to more parameters than the classical spatial scale and translation. These parameters are the spatio-temporal scaling and translation but also the rotation and the velocity. We finally obtain the **expression in the Euclidian, or Spatio-Temporal direct, domain, of the “motion-tuned” wavelet transform (MTSTWT)** with a set of tuning parameters  $g = \{a, c, \theta, \vec{b}, \tau\}$  :

$$\psi_{(a,c,\theta,\vec{b},\tau)}(\vec{x}, t) = a^{-3/2} \times \underbrace{e^{-\frac{c^{-2/3}}{2a^2} |\vec{x}-\vec{b}|^2} \times e^{-j\frac{c^{-1/3}}{a} \vec{k}_0 r^\theta (\vec{x}-\vec{b})}}_{\text{spatial term}} \times \underbrace{e^{-\frac{c^{4/3}}{2a^2} (t-\tau)^2} \times e^{-j\frac{c^{2/3}}{a} \omega_0 (t-\tau)}}_{\text{temporal term}} \quad (4)$$

**Remark** : A sequence is viewed as a spatio-temporal object and is decomposed on a basis tuned to 2D+T. Motion could be analyzed on 2D signals but we prefer to work on a 3D or, better called, “2D+T” approach because of a lack of consideration of the time variable in the 2D case.

## 2.1 Anisotropy parameter

An anisotropy parameter is used together with the velocity parameter in order to control the variance of the wavelet

w.r.t. to the velocity plane [11]. This parameter is introduced into the expression of the wavelet by replacing  $\vec{x}$  and  $\vec{k}$  by  $A^{-1}\vec{x}$  and  $A\vec{k}$ , respectively, with :

$$A = \begin{bmatrix} 1 & 0 \\ 0 & \epsilon \end{bmatrix} \quad (5)$$

and

$$A^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1/\epsilon \end{bmatrix} \quad (6)$$

The purpose of the anisotropy parameter, as explained in [11, 3], is to relax or increase the selectivity of the wavelet in speed. In the examples shown further, the selectivity has been fixed to 100. Adjusting this parameter can be used either for the “exhaustive” speed analysis of a scene or for the detection of an accurate speed value in a scene.

## 2.2 Wavelet and algorithm choice

The known initial authors [5, 8] of motion-tuned wavelets have been trying Morlet wavelets in the spectral domain.

We have investigated other wavelets like the partial gaussian derivatives. These wavelets have been used by Mallat et al., in the “dyadic transform” for singularity detection and multicontours decomposition/reconstruction. They are better suited to object detection in a scene because they oscillate much less than the Morlet wavelet. However, and maybe because of their difference in compacity between the direct and the Fourier domain, they have not given, for now, better results than the Morlet wavelet.

- Fast “A Trous” (“With Holes”) algorithm :

This algorithm is supposedly an interesting investigation field for the MTSTWT. We have built a motion-tuned transform based on this algorithm. But the wavelet used have to satisfy the AMR (Multi Resolution) condition (the twin-scale relation) which is not the case for the Morlet wavelet. Our present work is today concentrated on the construction of a filter bank satisfying the AMR condition (this is the case for Splines mentioned in [12]) and based on wavelets tuned to motion and more specifically to speed.

- Twin-scale relation for an AMR.

In order to satisfy to an AMR decomposition, the studied function, or signal, has to be projected on a “scaling” function. This function is related to the wavelet function, which, for the “a trous” algorithm gives the remaining information contained in the complementary orthogonal subspace [12]. The scaling function exists only for some classes of wavelets and must satisfy the following necessary condition called the “twin-scale relation” :

$$\frac{1}{2}\phi\left(\frac{x}{2}\right) = \sum_n h(n)\phi(x-n) \quad (7)$$

- Speed tuning admissibility conditions :

In order to make wavelets tuned to velocity, special conditions must be satisfied. These constraints come from the application of the velocity operator  $\Lambda^c$  on the wavelet (see equation 1). The  $\Lambda$  operator acts on the wavelet according to the following relation :

$$[\Lambda^c\psi](\vec{x}, t) = \psi(c^{-q}\vec{x}, c^p t) \quad (8)$$

The constraints result in two linear equations based, first on a unitary assumption and second on the application of the speed-tuning transformation to the  $v_0$  speed-plane [11]. They explicit the choice of the following  $p = 2/3$  and  $q = 1/3$  values.

### 3 Spectral and direct algorithms

We have tested a spectral domain and a direct domain algorithm for the MTSTWT. The direct version computes convolutions separably in the direct domain with the same filter length (-4,4, sizeof(image-direction)) as in the spectral domain. At this point we could argue that computing the MTSTWT in the direct domain is faster than going forth and back to the Fourier domain.

But the interesting point is of course the advantage in the direct domain to use a "compact" representation of the wavelet filter. Another interesting point is that the filter condition of compacity, in the direct domain, is easily satisfied in comparison with the spectral algorithm where a representation compact in the time domain as well as in the frequency domain is more difficult. This is a major concern when computing the WT in the Fourier domain, so this restriction, due to the Heisenberg principle, does not apply to the convolution computation in the direct spatio-temporal space.

The direct space, separable, version of the MTSTWT algorithm computed with the same filters lengths is slower than the spectral version. Nevertheless, in the direct space the computation should be faster under the condition to use fewer taps filters (compact filters). In the direct space the filters can also have a length inferior to the signal. In the Fourier space, they must have the same length for an implementation with term-to-term products.

- Other remark : The wavelets can be "scaled" by means of the parameter  $a$ . The change in variable is  $k/a$  for the spatial part and  $t/a$  for the temporal part. At this point we suggest not to tune the temporal variable to scale. The meaning of temporal variable scale tuning is not clearly understandable.

## 4 Results

### 1) Algorithm complexity of the MTSTWT

- The MTSTWT in the Fourier domain has a complexity of  $O(\text{filter\_length} \times (N^3 \text{Log}N))$  with  $N = m \times n \times k$  the size of the sequence.

- Add one IFFT3 per speed parameter if the analysis is made in the direct domain, or IFFT1 if the analysis is made in the spectral domain (not yet implemented). The speed analysis is based on the use of a set of wavelets tuned to different speeds. The speed analysis of a sequence is finally based on the search for a best basis in a dictionary of speed-tuned wavelets. The set of speed can be for example  $S_c = \{1, 3, 6, 12, 24, 48\}$  pixels/frame. This set can be chosen with or without an a priori knowledge of the speeds of the object pertaining the sequence. An a priori search in scene analysis would be to search only for objects having a speed around 3 pixels/frame. The term "around" has also the special meaning that an anisotropy parameter can be added to the wavelet in order to give it a specific variance. This anisotropy parameter, in other terms, gives the wavelet a selectivity. So for a wavelet tuned to 3 pixels/frame, the anisotropy  $\epsilon$  enables to search for speeds between, for example, 2 and 4 pixels/frame. The selection can finally be done by hard or soft thresholding (the so-called "keep or kill" and "shrink or kill" procedures).

- Add the FFT3 of the sequence-block analyzed prior to the MTSTWT (convolution products between block-sequence and spectral wavelet).

### 2) Computational speed

- The MTSTWT in the direct space with wave filters of length sizeof(each-image-direction), i.e. with the same filter lengths as in the Fourier domain, gives a computation speed of approximately 30 times (30 secondes) the computation in the spectral domain (see below for the spectral algorithm results). The computations have been done on a Xeon-BiProc at 2.6GHz.

- The MTSTWT in the spectral domain, with 3 different speed bases i.e. with 3 (2D+T)-wavelets tuned to 3 speeds (3, 6, 10 pixels/fr) on a  $360 \times 240 \times 8$  block of frames (Tisch-Tennis player sequence) takes 1200ms on the same Xeon bi-proc 2.4GHz. Add approximately 380 ms for the IFFT3 for each speed, which give a total of 2400ms, at one (the highest) resolution.

- The Fast Optical Flow [1] computed between 2 frames of the same sequence and at 4 resolutions takes 10 seconds on the 2.4GHz Xeon .

## 5 Conclusion and futur developments

Motion-tuned wavelets are very efficient for motion parameter computation, can be very accurate, robust to noise and occlusion due to their redundancy, and have the property of scalability by construction (multiresolution). Their application to compression is not obvious directly for motion

computation because :

- 1) A good segmentation of objects is difficult and not yet well-performed especially in MPEG4.
- 2) The MTSTWT computation is complex : they are similar in a way to matching pursuit due to the "large parameter sweep" that has to be done for each GOF (Group Of Frames), but we are working on the improvement of the algorithm.

On another side :

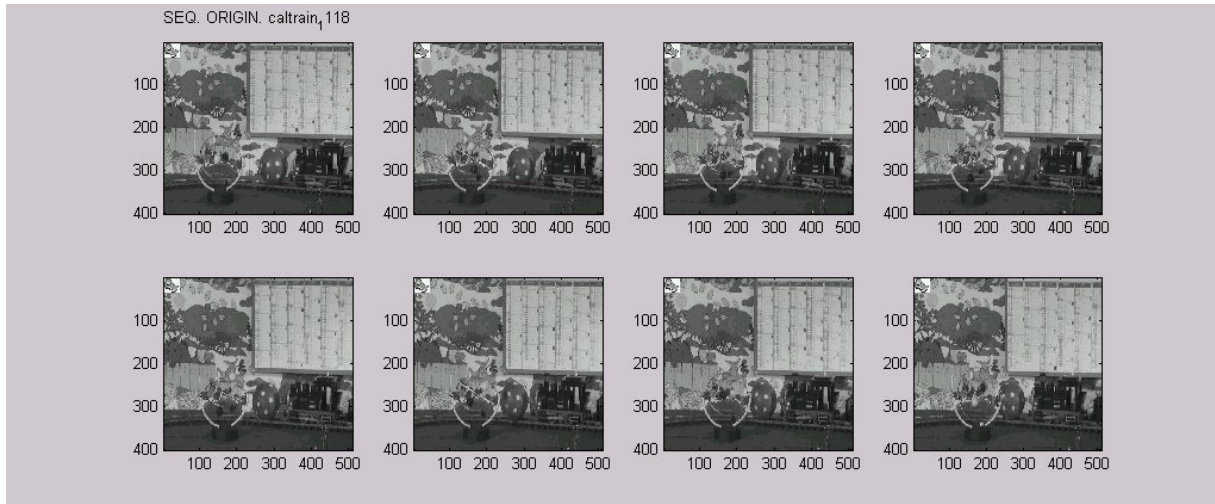
- 1) They can be a good solution in post motion-compensation trajectory filtering, either on a dense field or on block trajectories.
- 2) Their efficiency in motion parameter extraction can be used in scene analysis.
- 3) They have shown good results in target tracking.

Our present improvements are concentrated on a several precise points :

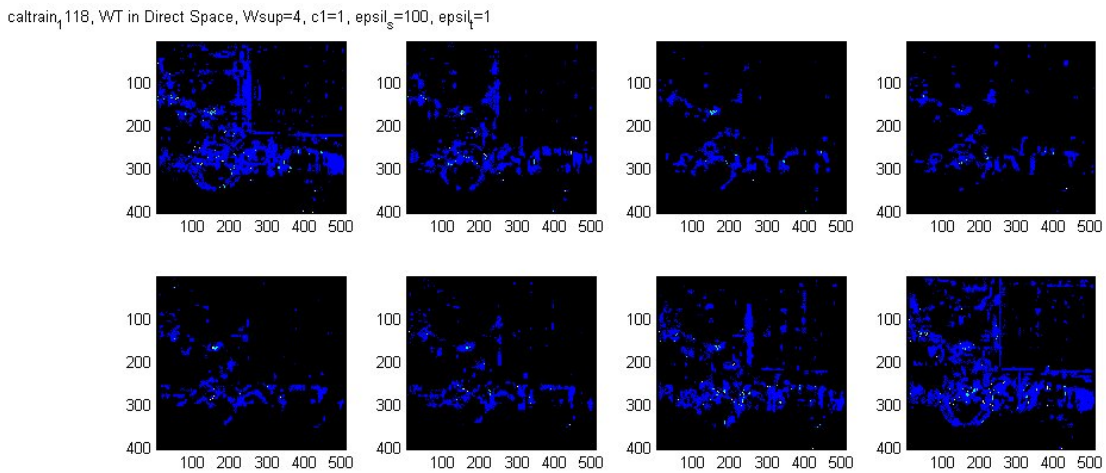
- 1) The use of wavelets better suited to object detection and less oscillating (Gaussian derivatives)
- 2) A computation in the direct space (discussed above) in order to avoid the direct to dual space conversions ( $FFT \rightleftharpoons IFFT$ )
- 3) The reduction of the convolution kernel size (especially for the direct domain computation)
- 4) The use of a Fast Algorithm based on the "A trous algorithm" and/or computation in direct space (Euclidian space convolution; already coded)
- 5) The extension of the wavelets tuning to acceleration and objects deformability [9, 4]

## References

- [1] Bernard C., Wavelets and ill-posed problems : the optical flow measurement and the irregular interpolation with several variables, PhD Thesis, Ecole Polytechnique, France, nov. 1999.
- [2] Brault P., "Motion Estimation and Video Compression with Spatio-Temporal Motion-Tuned Wavelets", *WSEAS Multi Conference on Applied Mathematics* (to be published in *WSEAS Transactions* ), Malta, september 2003.
- [3] Brault P., A New Scheme For Object-Oriented Video Compression And Scene Analysis, Based On Motion Tuned Spatio-Temporal Wavelet Family And Trajectory Identification, *IEEE-ISSPIT03* , Darmstadt, dec. 2003.
- [4] Corbett J., Leduc J.-P., and Kong M., Analysis of deformational transformations with spatio-temporal continuous wavelet transforms, in *Proceedings of IEEE-ICASSP*, Phoenix Arizona, p.4, 1999.
- [5] Duval-Destin M. and Murenzi R., Spatio-temporal wavelets: Applications to the analysis of moving patterns, *Progress in Wavelet Analysis and Applications*, Editions Frontières, Gif-sur-Yvette, France, pp.399-408, 1993.
- [6] M. Holschneider, R. Kronland-Martinet, J. Morlet and P.Tchamitchian, "The à trous Algorithm", *CPT-88/P.2215*, Berlin, pp 1-22, 1988.
- [7] Hsiang S.T. and Woods J.W., Highly-Scalable and Perceptually Tuned Embedded Subband/Wavelet Image Coding, Rensselaer Polytechnic Institute, N.Y.
- [8] Leduc J.-P., Spatio-Temporal Wavelet Transforms for Digital Signal Analysis, *Signal Processing* 60, Elsevier, pp. 23-41, 1997.
- [9] J.-P. Leduc, J. Corbett, M. Kong, V.M. Wickerhauser and B.K. Ghosh, Accelerated spatio-temporal wavelet transforms: An iterative trajectory estimation, *ICASSP5*, pp.2777-2780, 1998.
- [10] J.-P.Leduc, F.Mujica, R.Murenzi, and M.J.T. Smith, Spatiotemporal Wavelets: A Group-Theoretic Construction For Motion Estimation And Tracking . *Siam J.Appl.Math . Society for Industrial and Applied Mathematics* Vol.61,No.2,pp.596-632. 2000
- [11] F. Mujica, J-P. Leduc, R. Murenzi and M.J.T. Smith, A New Motion Parameter Estimation Algorithm Based on the Continuous Wavelet Transform *IEEE Transactions on Image Processing*, Vol. 9, No 5., pp 873-888, 2000.
- [12] Starck J.L. Murtagh F. and Bijaoui A., *Image Processing and Data Analysis, The Multiscale Approach*, Cambridge University Press, 1998, reprinted 2000.

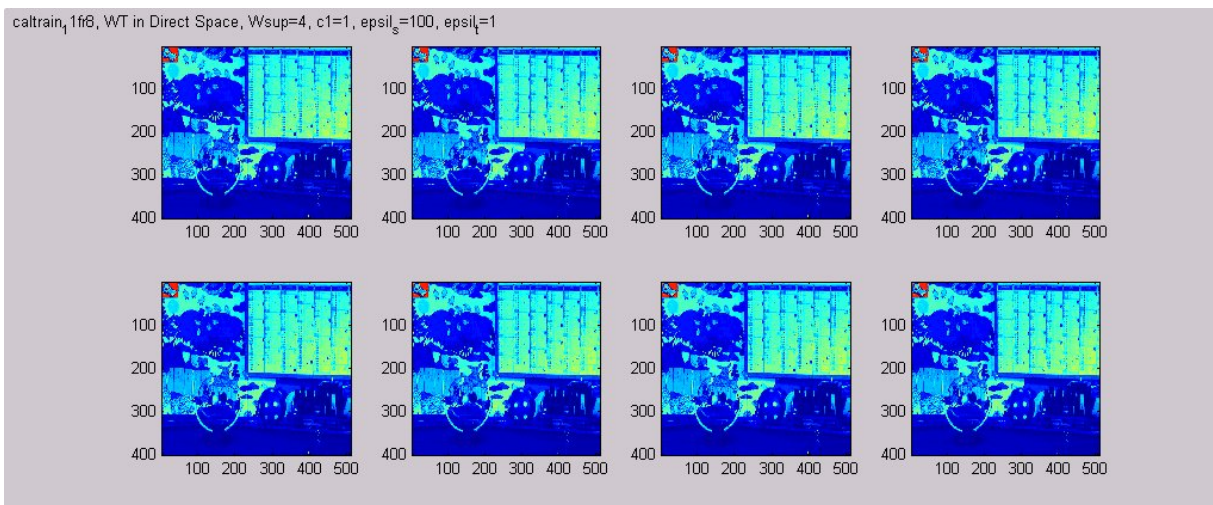


**Figure 1.** The “caltrain sequence” GOF used:  $400 \times 512 \times 8$ . In this sequence block of 8 frames, the train motion (and ball) is accelerated. All the objects (and background ) are moving : camera constant speed panoramic to the left to follow the train, calendar in vertical translation, two-balls pendulum with complex rotating motion etc.



**Figure 2.** MTSTWT analysis with the spectral algorithm and parameters  $a = 1$ ,  $\epsilon_s = 100$ , on 8 frames of the “Caltrain sequence”. Three interesting points to notice : 1) the first frame undergoes a “side effect” due to the non-circularity of the convolution (or non symmetrization or padding) 2) the accelerated motion can be detected by stronger (brighter coefficients) when getting closer to the 8th frame 3) All the motions are detected with strong (bright) or weak coefficients depending on the speed. In particular the upper left “thumb” image with a duck (logo of the software used for avi video decomposition into jpeg images) is totally invisible.





**Figure 3.** MTSTWT analysis with the spectral algorithm and parameters  $a = 1$ ,  $\epsilon_s = 100$ , on an “8 frames synthetic still-block” of the “Caltrain sequence”. This block of eight frames is synthetically composed of the same frame, i.e. there is no motion at all in these 8 frames. We can easily see, in comparison with the previous figure, that the MTSTWT has detected no motion. We can easily see on the upper left part of the image the “still image” logo of the software used for avi video decomposition