

Motion Estimation and Video Compression with Spatio-Temporal Motion-Tuned Wavelets

BRAULT P.

IEF, Institut d'Électronique Fondamentale, CNRS UMR 8622.

Université Paris-Sud

Bâtiment 220, 91405 Orsay Cedex

FRANCE.

patrice.brault@ief.u-psud.fr

Abstract : - We investigate a new scheme for hybrid video compression in the framework of an object-oriented approach and mainly based on the use of a motion-tuned wavelet transform. In the first step of this object approach, the wavelet transform is dedicated to the acquisition of the motion parameters like, commonly, the scale and the translation, but also the rotation, the shear and the object cinematic parameters like the speed and the acceleration. In the second step we base our scheme onto the construction of objects trajectories from the parameters acquired in the first step, and from a model chosen for the trajectory estimation. The third and final step consists in the object motion estimation based on the object trajectory built in the previous step.

So we can resume our scheme into three main steps: object motion parameterizing with wavelets, fast piecewise trajectory identification and motion prediction. The scheme proposed here is based on the notion of object trajectory which can be seen as an extension of the MPEG4 object and fixed block-matching approach and the H26L variable block-shape and long-time dependence approach.

Key-Words:- Continuous wavelet transform, spatio-temporal motion-tuned wavelets, object trajectory identification, motion estimation, video compression.

1 Introduction

We recall briefly here the concurrent notions of hybrid and 3D coders, the notion of spatial and temporal redundancy and the main methods which are used in their reduction. Other important aspects of video compression are also pointed out. We discuss the use of wavelets in video coding, make a short state of the art and argue for the choice of a wavelet family tuned to motion (as an alternative to optical flow constructions), a family quite different to orthogonal basis. We complete our scheme with the parameter computation of objects motions, with the trajectory identification and with the motion prediction associated to the iden-

tification.

2 Block-based matching strategy for motion estimation.

Motion estimation in the first generation video coding schemes such as H.261 aims at predicting a frame from its predecessors in order to reduce temporal redundancy. This reduction of temporal redundancy associated to a reduction of the spatial redundancy (e.g. by DCT, WT, K.L. or Hadamard transforms) make the "hybrid" coder. On the other side we have 3D coders which treat the video sequence as a whole 3D object and code, with motion estimation (lifting scheme) or

without, the sequence with a 3D or “2D+T” transform (e.g. WT in a separable or non-separable version).

In the case of hybrid coders and of temporal redundancy search, the true motion field is less important than the apparent motion field which must take illumination changes and noise into account to produce the lowest possible prediction error. The apparent motion field is usually found by a simple matching strategy in which the current frame is partitioned into blocks and each block is assigned a motion vector which identifies the best matching block in the previous frame according to a mean intensity difference criterion [13].

3 Object Oriented approach for M.E.

3.1 Introduction.

We start from the idea that a block-motion strategy, in object approaches, has a weaker interest due to the fact that frames are already segmented into objects and include a supplementary information for motion estimation than the crude partitioning into blocks. Motion estimation should now take into account the specific characteristics of the object which are its specific shape (contour), its specific intensity and colour, and its specific motion.

Shape recognition can closely be associated to motion modeling for most of the affine, scale, translation, rotation and shear can help recognize the object itself in any position.

Added to these, the kinematic parameters, i.e. velocity and n^{th} order acceleration can also help recognize the object; segmentation from motion [5, 6, 10, 14, 16]. But the kinematic parameters are intended to help modeling the **trajectory** of and object and predict its motion.

3.2 Motion detection, modeling and trajectory prediction.

In the framework of any object-oriented compression like MPEG4 V.2, where the segmentation and texture compression is standardized, our idea is to track each object of interest by first computing its kinematic parameters and then computing its trajectory model. The idea of assigning each object a trajectory is to realize a

motion prediction for each object. If we compare this approach with the block-matching approach still used in object-oriented MPEG4, we have a totally different view of considering the objects.

In H26L and H264, the compression schemes are not object-oriented. The temporal redundancy reduction is based on block-matching but the blocks can be clustered in different ways. The 8x8 or 16x16 elementary blocks can be grouped to create larger blocks of any size and the frame can even be split in horizontal or vertical bands.

So not only the coding of the object is different but also its motion prediction. Knowing the displacement of an object throughout a few frames, we build a trajectory and start predicting the whole object displacement based on the first trajectory points. The difference between the predicted trajectory and the real one provides the difference between the predicted object position and its true position, the last one being given in an “intra frame”. This difference gives an error that makes possible the trajectory control by a closed loop. This procedure has been in part inspired by the fact that “sprite” parts of scenes are associated a camera motion transmitted with the remainder of the quickly moving objects. The camera motion is measured and known at the source and the sprite is a “panoramic” image of the background that is taken at the very start of the sequence and transmitted only once to the decoder. Then each following frame of the same sequence is reconstructed from this initial panoramic image plus the motion parameters of the camera transmitted for each frame. This way only a few motion parameters for the camera are transmitted thus reducing drastically the bitstream to the decoder.

Our approach resembles this one in the way that we want to compute and predict the motion parameters of the objects which are not known initially. So the difficult task is to compute these parameters for one or several objects of the scene (the “intra”, i.e. present, image). An adaptation to specific applications can be to compute the motion parameters for objects pertaining a particular range of speed for example. This procedure can lead us to gather the objects by range of speed, of acceleration, of rotation ...

This procedure can be repeated in parallel for a group of objects of interest and especially for objects having close motion parameters, i.e. similar velocities. In this

framework, objects can be classified by range of velocity, range of orientation, range of acceleration a.s.o.

3.3 Motion detection and modeling with a wavelet transform.

A Motion Tuned Spatio-Temporal Wavelet Transform (MTSTWT) is at the basis of the motion parameters acquisition. This transform is a redundant transform (CWT) which can offer a good robustness in the case of object occlusions. This is also a multiresolution transform which is of interest for coding/decoding scalability.

Instead of making a frame, or object, warping to find the transformation (motion) parameters, the wavelet is tuned to a specific motion and no warping is done. If we take the case of velocity, a strong coefficient for a c_0 tuned wavelet will indicate that the object moves at a speed c_0 . In order to find velocities in a certain range, e.g. from 9 to 10 pixels/frame, we will use a high selectivity coefficient. This selectivity is also called "anisotropy", for it compresses the wavelet more or less along the chosen parameter (velocity). This anisotropy parameter is further developed in the following.

4 Fourier transform of a moving object.

The speed detection with wavelets is based on a property of the Fourier transform. The spatial spectrum of a N-Dim signal, or the object it represents, is shifted, in the Fourier domain, along the temporal wave-vector (i.e. the pulsation) by a value equal to the temporal velocity of the signal.

The relationship between the energy distribution of a static object $s(\vec{x}, t) = s(\vec{x})$ and the energy of the same object in linear motion $s(\vec{x} - \vec{v}t, t)$ can be mathematically represented by Fourier pairs. For a static object we have:

$$s(\vec{x}, t) = s(\vec{x}) \leftrightarrow \hat{s}(\vec{k})\delta(\omega) \quad (1)$$

and for an object in linear motion, so if we make the variable change $\xi = (\vec{x} - \vec{v}t)$, we have:

$$s(\vec{x} - \vec{v}t, t) = s(\vec{x}, 0) \leftrightarrow \hat{s}(\vec{k}, \omega + \vec{k} \cdot \vec{v}) \quad (2)$$

where the term $s(\vec{x}, 0)$ represents the signal at initial time 0, so the signal unwarped from its position at time t. The term $\vec{v}t$ represents a timely constant translation in the position space \vec{x} .

The equations 1 and 2 are representative of the redistribution of the energy between an object without motion and the same object having a linear motion. In the case of equation , the energy is concentrated in the null frequencies plane, $\omega = 0$. In the presence of a linear motion, this energy is distributed over a plane defined by:

$$\vec{k} \cdot \vec{v} + \omega = 0 \quad (3)$$

which can also be expressed:

$$\begin{bmatrix} \vec{k}' \omega \\ 1 \end{bmatrix} \begin{bmatrix} \vec{v}' \\ 1 \end{bmatrix} = 0 \quad (4)$$

Here the prime symbol is the transpose operator.

The relationship 4 defines a velocity plane orthogonal to the vector $[\vec{v}' \ 1]'$.

When the object presents an acceleration, the energy is spread around the dominant velocity plane. A fixed short time **linear approximation** is realized by using a time-window $w(t)$. This temporal windowing can be included in the velocity filter. The acceleration expressed in velocity/s can then be linked to the acceleration in velocity/frames times the number of frames per second.

4.1 Fourier transform of a moving signal or image.

If a signal or an image is moved in the space domain at some v_0 speed, the spectrum of the same static signal in the Fourier space, is shifted along the temporal wave-vector, by a value equal to the speed V_0 .

Example for a rectangle:

- 1) Static rectangle spectrum
- 2) Moving rectangle spectrum

4.2 Fourier transform of rectangles in translation.

We analyze through the Fourier transform, a sequence of 5 rectangles in translation. Three are in horizontal translation with speeds $V_h = 1, 3$ and 10 pix/fr and two are in vertical translation with speeds of $V_v = 1$ and 3 pix/fr.

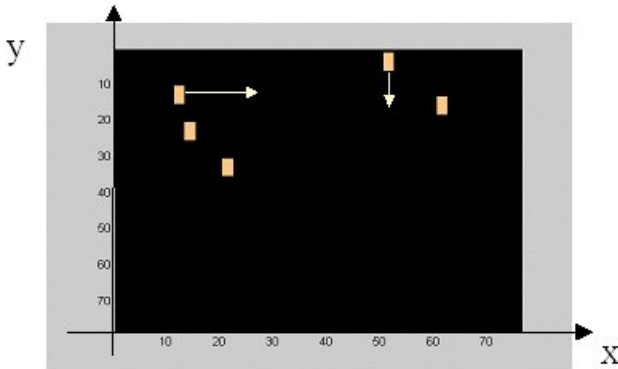


Figure 1. 16 frames sequence of rectangles in horizontal ($V_h = 1, 3$ and 10 pix/fr) and vertical ($V_v = 1$ and 3 pix/fr) translation.

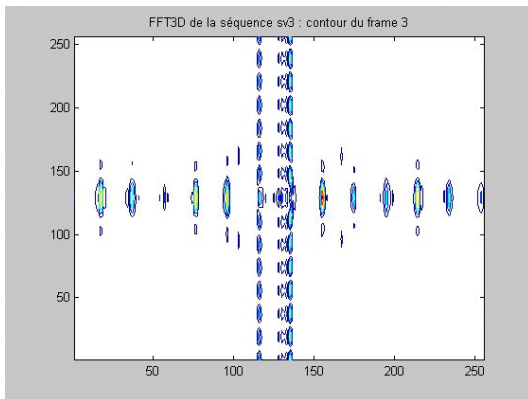


Figure 2. Wave-vector (k_x, k_y) spectrum of a moving rectangle (frame 3). The spatial spectrum exhibits the standard SINC shape spectrum (in directions x and y) of a rectangle.

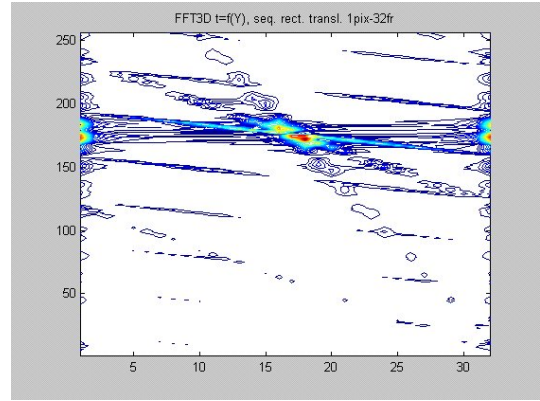


Figure 3. Wave-vector frequency (k_y, k_t) spectrum of 3 horizontally moving rectangles, at constant speed. The 3 slopes correspond to the 3 horizontal speeds.

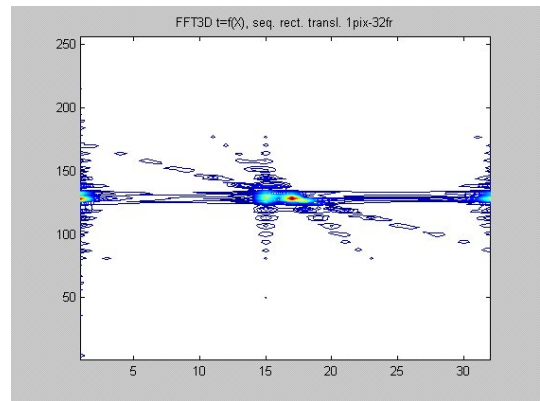


Figure 4. Wave-vector frequency (k_x, k_t) spectrum of 2 vertically moving rectangles, at constant speed.

5 Construction of the Motion-Tuned Spatio-Temporal Wavelet Transform (MTSTWT).

We have seen that the Fourier Transform applied in the 2D+T domain has the property to detect different velocities. Its main drawback is, similarly to any spectrum analysis, not to be able to give the position of a particular frequency, i.e. a velocity in this case. This is a capacity of the wavelet transform which provides the

frequency (scale) as well as the position (time) value of this velocity. We thus are interested in wavelets tuned to the 2D+T domain. More generally, we are interested in a wavelet that can be tuned to a specific kind of basic transformation (scale, translation, but also speed, acceleration, shear) to acquire motion parameters.

Here starts the construction of the of the Spatio-Temporal Motion-Tuned Wavelet Transform which is described here in a few steps:

- 1) Transfer of the motion operator from the signal to the wavelet.

This means that the signal is not motion-unwarped, and then the motion parameters are not computed by unwarping, but the wavelet is tuned to specific motion parameters (e.g. speed, rotation) and the signal is analyzed with this tuned wavelet (see [8, 11, 12]). The motion parameters are extracted from the wavelet coefficient.

- 2) Motion modelization and choice of the associated operators (or transformations).

We mean here that a set of parameters, so a limited motion models, are used. They are the translation, the scale, the rotation, the speed and can eventually be the acceleration and the deformation (shear).

- 3) The set of transforms applied to the wavelet is expressed under the shape of a “composite” transform.

- 4) Choice of the “mother wavelet”: the classical spatio-temporal Morlet wavelet, usually tuned only to spatial translation and scale, is used. This wavelet is a good candidate because it has the property of compactness in time and frequency. Nevertheless this wavelet has one drawback which is to be too oscillating. Subsequent work will probably make use of other wavelets like the B3-spline. The composite transform applied to a B-spline is under process.

- 5) Application of the composite transform mentioned above to the mother (here Morlet) spatio-temporal wavelet. This step makes the wavelet “motion-tuned”.

- 6) The final Transform is finally obtained by the convolution {wavelet \otimes video sequence} in the wave-vector frequency domain.

5.1 Motion Transforms in the Spatio-temporal approach (Galilean and DDM wavelets).

We first suppose that the mother wavelet support is concentrated in a velocity plane defined by:

$$\omega = -k\vec{v}_0.$$

The transformations used for a derivation of the set of bases for the CWT (2+1)D are:

- **Spatial and Temporal Translation** $T^{(\vec{b},\tau)}$

$$[T^{(\vec{b},\tau)}\psi](\vec{x}, t) = \psi(\vec{x} - \vec{b}, t - \tau) \quad (5)$$

$$[T^{(\vec{b},\tau)}\hat{\psi}](\vec{k}, \omega) = e^{-j(\vec{k}\cdot\vec{b} + \omega\tau)} \cdot \hat{\psi}(\vec{k}, \omega) \quad (6)$$

- **Rotation**

The R^θ transformation realizes a rotation of the wavelet on the spatial coordinates around the frequency axis:

$$[R^\theta \psi](\vec{x}, t) = \psi(r^{-\theta}\vec{x}, t) \quad (7)$$

$$[R^\theta \hat{\psi}](\vec{k}, \omega) = \hat{\psi}(r^{-\theta}\vec{k}, \omega) \quad (8)$$

with

$$r^{-\theta} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \quad (9)$$

- **Change in scale**

$$[D^a \psi](\vec{x}, t) = a^{-3/2} \psi\left(\frac{\vec{x}}{a}, \frac{t}{a}\right) \quad (10)$$

$$[D^a \hat{\psi}](\vec{k}, \omega) = a^{3/2} \hat{\psi}(a\vec{k}, a\omega) \quad (11)$$

- **Speed tuning**

This transformation can be considered as two change of scale operations on the spatial and temporal variables. This enables a localization of the wavelet around a velocity plane which has the correct inclination. This will be explained in an introduction on the property of the Fourier transform, where a spatial shift in the direct domain (i.e. the spatio-temporal domain) and with a specific speed, is equivalent to a shift in the Fourier domain (k_x, k_y, k_t) of the spatial spectrum along the temporal wave-vector (k_t).

$$[\Lambda^c \psi](\vec{x}, t) = \psi(c^{-1/3}\vec{x}, c^{2/3}t) \quad (12)$$

$$[\Lambda^c \hat{\psi}](\vec{k}, \omega) = \hat{\psi}(c^{+1/3}\vec{k}, c^{-2/3}\omega) \quad (13)$$

with example values of: $c = 1, 3, 10$ pixels/frame. Knowing the frame rate, the frame size and the object dimension, we can easily determine the object real time speed.

5.2 A short explanation: significance of the motion tuning.

The transformation which enables the tuning of a wavelet to velocity is obtained simultaneously by contraction/dilation in the position space and respectively dilation/contraction in the time space [8]. In the combination of these operations, the wavelet volume is maintained. The transformation can be expressed:

$$\Lambda^c \psi(\vec{x}, t) = \psi(\Lambda(\vec{x}, t)) \quad (14)$$

with:

$$\Lambda(\vec{x}, t) = (c^{1/3}\vec{x}, c^{-2/3}t) \quad (15)$$

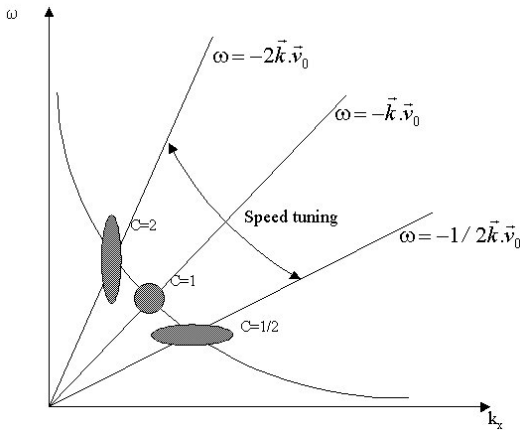


Figure 5. Speed tuning: displacement and distortion of the wavelets along hyperboles.

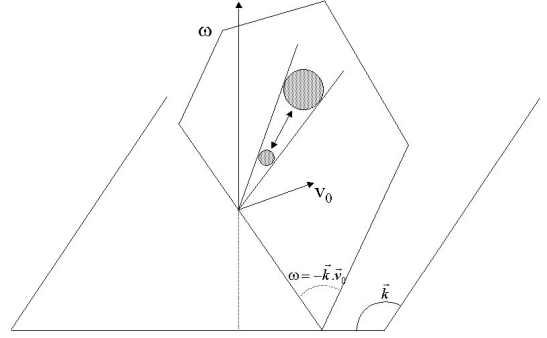


Figure 6. Cone of concentration of the wavelets around a constant speed plane given by $\omega = k\vec{v}_0$

5.3 Wavelet choice.

We now concentrate on the wavelet itself. The Morlet wavelet is a good candidate for it owns the properties of compactness in time and frequency, as seen before, which offers the possibility to realize computations in the Fourier space while keeping a good accuracy in the temporal speed domain. This is a complex valued wavelet. The 1D version in the spatial direct space can be expressed under the form of the product of a gaussian with a complex exponential of frequency k_0 :

$$\psi_{k_0}(x) = e^{-\frac{1}{2}x^2} \cdot e^{ik_0x} \quad (16)$$

- The **spatio-temporal** version of the classical Morlet wavelet is expressed under the form of modulated gaussians in the directions x, y, t :

$$\psi(x, t) = (e^{-1/2|\vec{x}|^2} \cdot e^{ik_0x} - e^{-1/2(|\vec{x}|^2 + |k_0|^2)}) \times (e^{-1/2t^2} \cdot e^{i\omega_0 t} - e^{-1/2(t^2 + \omega_0^2)}) \quad (17)$$

that we can rewrite:

$$\psi(x, t) = \underbrace{e^{-\frac{x^2+t^2}{2}} \cdot e^{-i(k_0x + \omega_0 t)}}_A - \underbrace{e^{-\frac{x^2+t^2}{2}} \cdot e^{-\frac{k_0^2 + \omega_0^2}{2}}}_B \quad (18)$$

This last expression can be limited to the A term if the admissibility term B is negligible. In fact this happens

for $|k_0|$ and $\omega_0 \geq \pi \sqrt{\frac{2}{Ln2}} \simeq 5.336$. In the following we use this same value for $|k_0|$ and ω_0 , which leads us to a speed: $v_0 = \frac{\omega}{k} = 1(\text{pix}/\text{fr})$.

We thus obtain the expression of the **simplified spatio-temporal Morlet wavelet**, by cancellation of the admissibility term:

$$\psi(x, t) = (e^{-1/2|\vec{x}|^2} \cdot e^{ik_0x}) \times (e^{-1/2t^2} \cdot e^{i\omega_0t}) \quad (19)$$

- The wave-vector version of the Morlet wavelet in the reciprocal frequency (Fourier) space is:

$$\hat{\psi}_{k_0}(k) = e^{\frac{1}{2}(k-k_0)^2} \quad (20)$$

and its spatio-temporal version in the same space is:

$$\hat{\psi}(\vec{k}, \omega) = \left(e^{-\frac{1}{2}|k-k_0|^2} - e^{-\frac{1}{2}(|k|^2+|k_0|^2)} \right) \times \left(e^{-\frac{1}{2}(\omega-\omega_0)^2} - e^{-\frac{1}{2}(\omega^2+\omega_0^2)} \right) \quad (21)$$

Like for the direct space we will use a simplified version of the Morlet wavelet by cancelling the admissibility term, which will give the "simplified Spatio-Temporal spectral" version of the Morlet wavelet.

$$\hat{\psi}(\vec{k}, \omega) = \left(e^{-\frac{1}{2}|k-k_0|^2} \right) \times \left(e^{-\frac{1}{2}(\omega-\omega_0)^2} \right) \quad (22)$$

Remark: if k_0 and ω_0 are large enough, i.e. we can remove the admissibility terms, the expressions in both direct and Fourier domains are equivalent to a modulated gaussian or a Gabor filter.

5.4 The spatio-temporal composite transform.

We now look for the definition of a wavelet transform tuned to motion. We believe the affine transform, with speed tuning is a good candidate and is able to detect motions not only for their spatial trajectories but also for their kinetic property: the speed, or even more the acceleration. We will first concentrate on the simplest approach with only the velocity parameter.

The parameters set for this transform is given by:

- 1) the spatial homothety (scale parameter).
- 2) the rotation.
- 3) the spatial and temporal position.
- 4) the velocity.

We define [15] the Spatio-temporal continuous wavelet transform (STCWT) like the application of the finite energy signals from $\mathbb{R}^2 \times \mathbb{R}$ to:

$G = g : (a, c, \theta, \vec{b}, \tau) \in (\mathbb{R}^+ \times \mathbb{R}^+[0, 2\pi] \times \mathbb{R}^2 \times \mathbb{R})$, where the respective parameters of scale, velocity, rotation, spatial and temporal position, correspond to the parameters of the Galilean approach quoted in the paragraph 3.1.

The **composite transform** Ω_g defined by the application of all the operators on the wavelet, is expressed:

$$[\Omega_g\psi](\vec{x}, t) = [T^{\vec{b}, \tau} R^\theta \Lambda^c D^a \psi](\vec{x}, t) \quad (23)$$

and when replacing the transforms by their expressions:

$$[\Omega_g\psi](\vec{x}, t) = a^{-3/2} \psi \left(\frac{c^{-1/3}}{a} r^{-\theta} (\vec{x} - \vec{b}), \frac{c^{2/3}}{a} (t - \tau) \right) \quad (24)$$

In the same way, for the Fourier space:

$$\begin{aligned} [\Omega_g\hat{\psi}](\vec{k}, \omega) &= [T^{\vec{b}, \tau} R^\theta \Lambda^c D^a \psi](\vec{k}, \omega) \quad (25) \\ &= a^{3/2} \hat{\psi} \left(ac^{1/3} r^{-\theta} \vec{k}, \frac{1}{ac^{2/3}} \omega \right) e^{-j(\vec{k}\vec{b} + \omega\tau)} \quad (26) \end{aligned}$$

5.5 Composite transform applied to the Morlet wavelet.

The composite transform as expressed before requires now the choice of a wavelet. We have seen that the spatio-temporal Morlet wavelet tuned to a motion exhibits the characteristics of temporal and spatial compactness as well as admissibility. If we then replace the expression of the space and time variables modified by the **composite transform** (24) in the expression of the **simplified spatio-temporal Morlet wavelet** (19), we obtain:

$$\psi_{(a,c,\theta,\vec{b},\tau)}(\vec{x},t) = a^{-3/2} \times \underbrace{e^{-\frac{c^{-2/3}}{2a^2}|\vec{x}-\vec{b}|^2} \times e^{-j\frac{c^{-1/3}}{a}\vec{k}_0 r^\theta(\vec{x}-\vec{b})}}_{\text{spatial term}} \times \underbrace{e^{-\frac{c^{4/3}}{2a^2}(t-\tau)^2} \times e^{-j\frac{c^{2/3}}{a}\omega_0(t-\tau)}}_{\text{temporal term}} \quad (27)$$

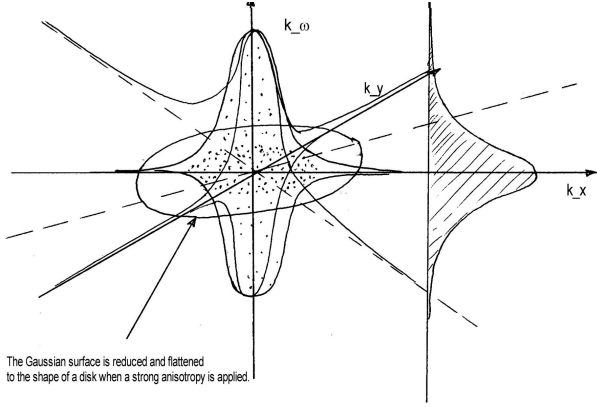


Figure 7. The spatio-temporal Morlet wavelet is representable by a 2D spatial gaussian surface (left). We obtain its 3D representation by varying the points density on this surface, along the vertical axis ω , by a frequential (or temporal) gaussian, represented on the right. The wavelet then takes the shape of the ellipsoid seen on the figure 5. When the wavelet is built with a strong temporal anisotropy ϵ_t , it is comparable to a disk (center of the left figure).

For we work with the separable version of this 3D filter, each filter can be represented like one Morlet wavelet. The speed tuning, as we said, is realized by contraction and dilation in the spatial space, respectively dilation or contraction in the temporal space. This is put into evidence on the figures below, where the wavelet is tuned to a velocity $c = 2$. The wavelet angular frequency, or wave vector, is chosen equal to $k_0 = 6$, which, as we saw formerly, enables to neglect the admissibility term in the wavelet model. The anisotropy parameter ϵ_s or ϵ_t (spatial or temporal) is introduced in the expression of the wavelet as a

multiplier to the contraction. If we take into account the speed tuning case, the purpose of the anisotropy parameter is to restrict or expand the domain of speeds that can be detected by one tuned wavelet, i.e. if we have a wavelet tuned to $c = 10$, a high anisotropy parameter will restrict the recognition to objects having a speed very close to $c = 10$. On the contrary, in the case of an anisotropy parameter equal to 1, the speeds recognized can have a larger range, with wave coefficients decreasing from the speed $c = 10$ to speeds of 1, 3, 20, etc. and a maximum at $c = 10$.

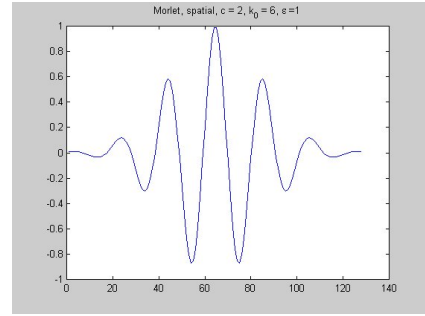


Figure 8. Morlet wavelet on the spatial axis, for $c = 2$ and $\epsilon_s = 1$

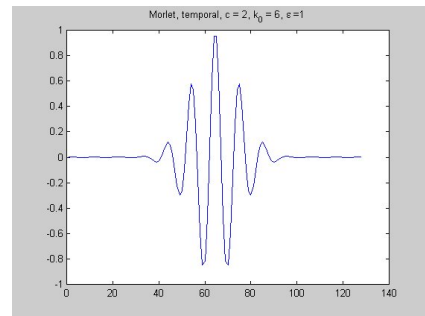


Figure 9. Morlet wavelet on the temporal axis, for $c = 2$ and $\epsilon_t = 1$. This shows the contraction along the temporal axis.

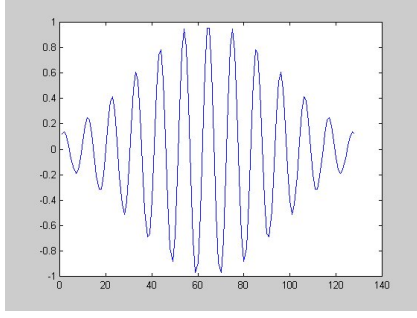


Figure 10. Morlet wavelet on the temporal axis, for $c = 2$ and $\epsilon_t = 10$

6 Specificity of the velocity tuning case.

Both computation in the direct and the spectral domain have been tried. Here we provide the expression for a computation in the Fourier domain which has the advantage of an easy computation (product terms) following a 3D FFT of each signal: mother wavelet and sequence. We then come back in the direct domain by an inverse FFT, but the sequence can be analyzed in the Fourier domain. Literally, the expression of the simplified spatio-temporal velocity-tuned Morlet wavelet in the Fourier domain is finally given by (see 13 + 19):

$$\hat{\psi}(\vec{k}, \omega) = \left(e^{-\frac{1}{2}(c^{1/3})^2|k-k_0|^2} \right) \times \left(e^{-\frac{1}{2}(c^{-2/3})^2(\omega-\omega_0)^2} \right) \quad (28)$$

We have the complete motion-tuned, Fourier space, simplified, ST, Morlet wavelet:

$$\hat{\psi}(\vec{k}, \omega) = a^{3/2} \left(e^{-\frac{1}{2}a^2(c^{1/3})^2|k-k_0|^2} \right) \times \left(e^{-\frac{1}{2}1/a^2(c^{-2/3})^2(\omega-\omega_0)^2} \right) \times \left(e^{-j(\vec{k}\vec{b}+\omega\tau)} \right) \quad (29)$$

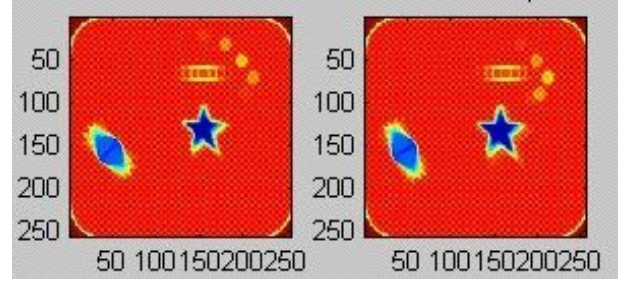


Figure 11. Two frames of a synthetic sequence analyzed by the MTSTWT

7 Energy densities and motion parameters extraction.

The energy density for each motion tuned filter is computed over the sequence. An energy peak in the velocity domain is representative of the presence of a particular velocity. This enables to establish a classification map of the encountered velocities in throughout the sequence.

The same procedure can be repeated for specific rotations, translations, changes of scale, shear.

8 Object trajectory identification and prediction

Once we have computed the spatial and kinematic parameters of objects, we identify in a quick process, see [9], and on the basis of a few frames, the trajectory of the object to a linear model. The parameters of the model can be the coefficients of an n-th order polynomial. A model under the form of an N-th order spline function can also be an appropriate choice as well as any N-th degree polynomial, assuming with have absolutely no knowledge of the object trajectory. We will consider in a second approach the possibility of having an a priori knowledge of the trajectory model.

9 Conclusion

We have posed here the building blocks of a new scheme for scene analysis and video compression. First we use a continuous wavelet transform with

mother wavelets tuned to motion. This provides an alternative method to optical flow computation and has demonstrated its efficiency in high speed targets tracking. The interest here is to realize a motion estimation scheme based on the static and kinetic parameters of the objects pertaining to a video scene. Then, in a second time, we propose to compute a model of the object trajectory based on polynomials or splines. This scheme integrates the motion in a few frames. Finally the object motion prediction replaces the classical BM for all or some objects of interest and the object trajectory is used for both motion compensation and scene analysis.

Acknowledgement

The author would like to thank Professor Alain Merigot, CNRS-Institut d'Electronique Fondamentale, for his helpful comments.

References

- [1] Adelson E. H. and Bergen J. R., Spatiotemporal Energy Models for the Perception of Vision, *J. Opt. Soc. Amer.*, vol. A2, pp 284-299, february 1985.
- [2] Barron J. L., Fleet D. J. and Beauchemin S. S., Performance of Optical Flow Technique, *IJCV International Journal of Computer Vision*, vol. 12, no. 1, pp 43-77, 1994.
- [3] Bernard C., Discrete Wavelet Analysis for Fast Optic Flow Computation, *Internal Report RI415 of the "Centre de Mathématiques Appliquées"*, École Polytechnique. February 26, 1999.
- [4] Brault P. and Mounier H., Automated, Transformation Invariant, Shape Recognition Through Wavelet Multiresolution, *SPIE01 International Society for Optical Engineering*, San Diego, USA, 2001.
- [5] Bonnaud L., Labit C. and Konrad J., Interpolative coding of image sequences using temporal linking of motion-based segmentation, *ICASSP95*, 1995.
- [6] M.M. Chang, A.M. Tekalp and M.I. Sezan, Simultaneous motion estimation and segmentation, *IEEE Trans. Image Processing*, vol.6(9), pp 1326-1333, 1997.
- [7] J.N.Driessen, L.Boroczky, and J.Biemon, Recursive Motion Field Estimation from Image Sequences, *J.Visual Commun.Image Reproduction*, vol.2, pp 259280, 1991.
- [8] Duval-Destin M. and Murenzi R., Spatio-temporal wavelets: Applications to the analysis of moving patterns, *Progress in Wavelet Analysis and Applications*, Y.Meyer and S. Roques eds., Editions Frontières, Gif-sur-Yvette, France, pp 399408, 1993.
- [9] M. Fliess and H. Sira-Ramirez, An Algebraic FrameWork For Linear Identification, to appear in *ESAIM Contr. Opt. Calc. Variat.*, Vol. 9, 2003.
- [10] S. Jehan, E. Debreuve, M. Barlaud and G. Aubert, Segmentation spatio-temporelle d'objets en mouvement dans une séquence vidéo par contours actifs déformables, *RFIA00*, Paris, 2000.
- [11] J.-P.Leduc, F.Mujica, R.Murenzi and M.J.T.Smith, Spatio-temporal wavelet transforms for motion tracking, *ICASSP97*, vol. 4, pp 30133016, 1997.
- [12] 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 596632, 2000.
- [13] Magarey J. and Kingsbury N., Motion estimation using a complex-valued wavelet transform, *IEEE Transactions on Signal Processing*, vol.46(4), pp 1069-1084, april 1998.
- [14] E. Memin and P. Perez, Dense estimation and object-based segmentation of the optical flow with robust techniques, *IEEE Trans. Image Processing*, 7(5), pp 703-719, may 1998.
- [15] F.Mujica, J-P. Leduc, R.Murenzi and M.J.T. Smith, Spatio-Temporal Continuous Wavelets Applied to Missile Warhead Detection and

- Tracking, *SPIE-VCIP 3024*, J.Biemond and E.J.Delp eds., Bellingham WA., pp 787 – 798, 1997.
- [16] J-M. Odobez and P. Bouthemy, Direct incremental model-based image motion segmentation for video analysis, *Signal Processing*, 66(2), pp 143-155, 1998.
- [17] B.Pesquet-Popescu and V. Bottreau, Three-dimensional lifting schemes for motion compensated video compression, *Proc. IEEE-ICASSP01*, Salt Lake City, 7-11 May 2001.
- [18] Wang J.Y.A. and Adelson E.H., Spatio-Temporal Segmentation of Video Data, *Proceedings of SPIE on Image and Video Processing II*, 2182, pp 120-131, San Jose, February 1994.
- [19] Weiss Y. and Adelson E.H. , Perceptually Organized EM: A Framework for Motion Segmentation that Combines Information about Form and Motion, report TR N 315, *MIT Media Lab Perceptual Computing Section*, 1994.
- [20] The H261 Standard
- [21] The MPEG2 Standard
- [22] The MPEG4 Standard
- [23] The H26L Standard
- [24] The H264 Standard