SPATIO-TEMPORAL WAVELET TRANSFORMS FOR MOTION TRACKING. *

Jean-Pierre Leduc   Fernando Mujica   Romain Murenzi   Mark Smith
Georgia Institute of Technology
Center of Signal and Image Processing
Clark Atlanta University
Center for Theoretical Studies of Physical Systems, Physics Department
Atlanta, Georgia

ABSTRACT

This paper addresses the problem of detecting and tracking moving objects in digital image sequences. The main goal is to detect and select mobile objects in a scene, construct the trajectories, and eventually reconstruct the target objects or their signatures. It is assumed that the image sequences are acquired from imaging sensors. The method is based on spatio-temporal continuous wavelet transforms, discretized for digital signal analysis. It turns out that the wavelet transform can be used efficiently in a Kalman filtering framework to perform detection and tracking. Several families of wavelets are considered for motion analysis according to the specific spatio-temporal transformation. Their construction is based on mechanical parameters describing uniform motion, translation, rotation, acceleration, and deformation. The main idea is that each kind of motion generates a specific signal transformation, which is analyzed by a suitable family of continuous wavelets. The analysis is therefore associated with a set of operators that describe the signal transformations at hand. These operators are then associated with a set of selectivity criteria. This leads to a set of filters that are tuned to the moving objects of interest.

1. INTRODUCTION

The primary purpose of the present work is to investigate families of spatio-temporal continuous wavelet transforms (CWT), and their utility for motion tracking and trajectory constructions. The approach considered in this paper differs fundamentally from other techniques that have been proposed such as those based on optical flow, pel-recursive, block matching and Bayesian models. The main novelty of this method is that it combines the CWT with Kalman filtering for tracking. Several families of CWTs can efficiently perform various tasks like motion-based detection and segmentation, selective tracking and reconstruction of objects in motion. The CWT is also highly robust against sensor noise. Moreover, it is able to handle temporary occlusions resulting from crossing trajectories. These properties are generally not found in techniques rooted in optical flow and block-based motion estimation.

The study of CWTs originally evolved from considering spatio-temporal affine transformations. These were easily amenable to Lie group structures and admissible wavelet representations. The approach turned out to be efficient for signal analysis and enabled the introduction of numerous physical parameters as criteria of selectivity. The importance of CWTs in this field was recognized several years ago [1]. Although the analysis of image sequences requires numerous analyzing parameters, only a small subset of them has to be considered simultaneously in each specific application. The most significant components of uniform motion are studied in this paper, i.e. the translation, the rotation, the deformation, and the acceleration. The spatial orientation (the preferential axis of inertia) and the scale are additional parameters of concern; indeed, the scale is intrinsic to any wavelet analysis. The application of motion tracking is addressed in this paper and is illustrated with CWTs tuned to the velocity, i.e. the translational motion. It is assumed that local motion is linear. Hence, the technique applies whenever the approximation is valid locally on a few frames (3 or more). CWTs that are tuned to velocities are called *Galilean wavelets* to refer to the Galilei group used in classical mechanics.

2. BUILDING FAMILIES OF CWTS

The construction of CWTs relies on signal transformations that model motion and object deformations. They can take into account translation, rotation, scale and shear. One idea developed in this work, consists of expressing all these elementary transformations as unitary operators in the spatio-temporal domain $2D + T$ (2D spatial plus time), and to write useful generalizations for uniform motion (i.e. motion described by time-invariant parameters), namely *translation*, *rotation*, *deformation*, and *acceleration*. These unitary operators and their related parameters are eventually combined to form a general law of signal transformation. This transformation is intended to be applied either to the signal or to the mother wavelets. When applied to the signal, it describes the transformations performed by the motion i.e. warpings of the signal. When applied to an admissible mother wavelet, it generates a whole family of continuous spatio-temporal wavelets (band pass filters for signal analysis). The construction of these families is ruled by locally compact groups (Lie groups) and the admissibility of a mother wavelet is enforced by three representation properties: square-integrability, irreducibility and unitarity. The procedure to calculate admissible wavelets is well-known and relates to that of the coherent states originating from theoretical physics. As such, it has already been developed in other research work in the one and multi-dimensional cases [2]. The construction proceeds as follows. Spatio-temporal transformations are first defined on the space $(\mathbf{R}^2 \times \mathbf{R})$. Typically, in this construction, the structure of the parameters leads to composition relationships, inverse and identity that characterize a group. The

*This material is based upon work supported in part by supported partly by the U.S. Army Research Office under grants DAAH-04-96-1-0161 and DAAH-04-95-1-0650, and in part by a Belgian NATO fellowship.

study of group representations in spatio-temporal Hilbert spaces comes thereafter. According to the theory of coherent states, the demonstration of unitarity, irreducibility and square-integrability of these representations guarantees the existence of admissible continuous spatio-temporal wavelets. Practically, this means that operating the mother wavelet in the space of the parameter definition covers the whole set of bandpass filters of finite energy, while preserving all the well-known wavelet properties (isometry, inversion, reproducing kernel, and resolution of the identity). In this section, five different constructions of CWT families will be considered as examples. They originate from the groups covering all the motions [3] and represent the action of Lie algebras and groups on manifolds. First, operators on wavelet and signal $\{\mathbf{\Omega} : L^2(\mathbf{R}^2 \times \mathbf{R})\} \rightarrow L^2(\mathbf{R}^2 \times \mathbf{R})\}$ will be defined with their respective set of parameters.

The affine-Galilei group supports the construction of CWTs tuned to velocity and *uniform translational motions.* The set of operators and parameters involved in this CWT are the spatio-temporal translation of parameter $\vec{b}$ and τ to represent the space and time locations, the velocity $\vec{v}$, the dilation a to represent the scale, and the spatial rotation θ to represent the orientation (the preferential anisotropy). The action of these parameters can be written as the following spatio-temporal transformation

$$\vec{x}_2 = \frac{1}{a} R(-\theta)(\vec{x}_1 - \vec{b} - \vec{v}t); \quad t_2 = t_1 - \tau \ , \tag{1}$$

where $R(\theta)$ is the rotation matrix in SO(2). Let us write the wavelet transform, $\Psi(\vec{x}, t)$, in the Galilean family. In the spatio-temporal domain, we have

$$\begin{aligned} &\left[\mathbf{\Omega}(\vec{b}, \tau, \vec{v}, a, \theta)\Psi\right](\vec{x}, t) \\ &= \tfrac{1}{a}\Psi\left[\tfrac{1}{a}\ R(-\theta)\ \left(\vec{x} - \vec{b} - \vec{v}t\right)\ ,\ t - \tau\right]\ , \end{aligned} \tag{2}$$

and in the Fourier domain, where $\vec{k}$ and ω stand for spatial and temporal frequencies, we have

$$\begin{aligned} &\left[\widehat{\mathbf{\Omega}}(\vec{b}, \tau, \vec{v}, a, \theta)\widehat{\Psi}\right]\left(\vec{k}, \omega\right) \\ &= a\ e^{-i\ (\vec{k}.\vec{b}\ +\ \omega\tau)}\ \widehat{\Psi}\left[a\ R(-\theta)\ \vec{k}\ ,\ \omega + \vec{k}\vec{v}\right]\ . \end{aligned} \tag{3}$$

The set of parameters considered in this family of CWTs is $(\vec{b}, \tau, \vec{v}, a, \theta)$. These CWTs are called Galilean wavelets.

A slightly different approach to Galilean wavelets, called the *kinematical wavelets*, has been described by Duval-Destin and Murenzi [1]. In this case, the set of parameters is $(\vec{b}, \tau, a, c, \theta)$ where a is the spatio-temporal dilation, and c is the speed parameter. c and θ reach the velocity. The spatio-temporal transformation is given

$$\vec{x}_2 = \frac{1}{c^{1/3}a} R(-\theta)(\vec{x}_1 - \vec{b}); \quad t_2 = \frac{c^{2/3}}{a}\ (t_1 - \tau) \tag{4}$$

Let us now consider *uniform rotational motion* as a third spatio-temporal transformation, and generate a family of CWTs. Uniform rotational motion is different from the spatial rotation of the SO(2) group in the sense that it incorporates time and space. The resulting velocity is given in this case by

$$\vec{v}(t) = \vec{v}_0 + \vec{\omega} \wedge \vec{x}(t)\ , \tag{5}$$

where $\vec{\omega}$ is the angular velocity, $\vec{v}$ is the translational velocity and $\vec{x}(t)$ the current coordinate location of the moving object. The symbol $\wedge$ stands for the cross vector product. Another way of expressing this signal transformation in the image planes is given as

$$\vec{x}_2 = R(-\theta t)\vec{x}_1; \quad t_2 = t_1 - \tau\ , \tag{6}$$

where $\left[R(\theta t) = \begin{pmatrix} \cos\theta t & -\sin\theta t \\ \sin\theta t & \cos\theta t \end{pmatrix}\right]$. The set of parameters considered in this CWT family is $(\vec{b}, \tau, a, \vec{v}_0, \vec{\omega})$ or $(\vec{b}, \tau, a, \vec{v}_0, \theta)$.

Uniform temporal dilation (i.e. expansion or contraction) is defined by substituting in Equation (6) $R(\theta t)$ by $\left[D(\alpha t) = \begin{pmatrix} e^{-\alpha t} & 0 \\ 0 & e^{-\alpha t} \end{pmatrix}\right]$. This transformation is important since any object in motion approaching the camera undergoes rather exponential expansions in the image field. The set of parameters of interest for the CWT construction are then $(\vec{b}, \tau, a, \vec{v}_0, \alpha)$.

A fifth set of analyzing parameters would consider *uniform acceleration* $\vec{\gamma}$, given by the second order coefficient when expanding the trajectory curve $\vec{x} = \vec{f}(t)$ in series

$$\vec{x}(t)\ =\ \vec{b}\ +\ \vec{v}_0 t\ +\ \frac{1}{2}\ \vec{\gamma}_0\ t^2\ +\ \sum_{n=1}^{\infty} \frac{1}{(n+2)!}\ \vec{\gamma}_n\ t^{n+2} \tag{7}$$

where $\vec{v}_0 = \frac{d\vec{f}(t)}{dt}|_{\vec{x}=0}$ is the velocity, and $\vec{\gamma}_0 = \frac{d^2\vec{f}(t)}{dt^2}|_{\vec{x}=0}$ is the acceleration. The $\vec{\gamma}_n$ stands for n^{th}-order acceleration and is not considered in this study. Thus, the parameters of interest in this CWT family will be $(\vec{b}, \tau, a, \vec{v}_0, \vec{\gamma}_0)$.

3. DEFINITION OF THE CWT

This section presents the definition of one CWT family, the *Galilean wavelets.* The signal $s(\vec{x}, t)$ subject to analysis is defined in the Hilbert space $L^2(\mathbf{R}^2 \times \mathbf{R}, d^2\vec{x}dt)$. The CWT $W[s; \vec{b}, \tau, \vec{v}, a, \theta]$ is defined as an inner product

$$\begin{aligned} W[s; \vec{b}, \tau, \vec{v}, a, \theta] &= c_\Psi^{-1/2}\ < \Psi_{\vec{b}, \tau, \vec{v}, a, \theta} \mid s > \\ &= c_\Psi^{-1/2} \int_{\mathbf{R}^2 \times \mathbf{R}} d^2\vec{x}dt \bar{\Psi}_{\vec{b}, \tau, \vec{v}, a, \theta}(\vec{x}, t)\, s(\vec{x}, t) \end{aligned}$$

where the overbar $^-$ stands for the complex conjugate. The wavelet, Ψ, is a *mother wavelet.* It must satisfy the condition of admissibility (i.e. of square-integrability) meaning that there exits a constant c_Ψ (normalized to one) such that

$$c_\Psi\ =\ (2\pi)^3\ \int_{\mathbf{R}^2 \times \mathbf{R}} d^2\vec{k}\ d\omega\ \frac{|\widehat{\Psi}(\vec{k}, \omega)|^2}{|\vec{k}|^2}\ <\ \infty\ .$$

A numerically efficient way of performing the CWT consists of working in the spectral domain by means of the (2D+T) FFT. The other CWT families have a similar definition.

4. EULER-LAGRANGE EQUATION

Let us consider Lagrange's principle of the least action that can be equivalently derived in classical mechanics and in optimal control from the calculus of variations. The system is characterized by the action S and a non-negative definite function, called the Lagrange function, $L[\vec{x}(t), \dot{\vec{x}}(t); t]$, where $\vec{x}(t)$ is the trajectory and $\dot{\vec{x}}(t) = \frac{d\vec{x}(t)}{dt}$ is the corresponding velocity function. The calculus of variations allows us to derive the motion equation and the trajectory that optimize the action. Usually, motion between times t_1 and t_2 in a conservative mechanical system coincide with the extremal of the functional

$$S = \int_{t_1}^{t_2} L[\vec{x}(t), \dot{\vec{x}}(t); t] dt \ , \tag{8}$$

where L is the difference between the kinetic and the potential energy. Optimal control exploits the same modeling, where S is a cost function to be optimized under some constraints to be specified. The trajectory is then uniquely defined when the initial conditions are known in terms of object location and velocity (detection issue). At the extremum, denoted by $*$, the calculus derives the well-known Euler-Lagrange equation

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\vec{x}}^*} - \frac{\partial L}{\partial \vec{x}^*} = 0 \ . \tag{9}$$

In this paper, the Lagrange function L to be considered is the square of the modulus of the Galilean CWT, i.e. the energy density $| < \Psi_{\vec{b},\tau,\vec{v},a,\theta} \mid s > |^2$, $\vec{b} = \vec{x}$, $\tau = t$ and $\vec{v} = \dot{\vec{x}}(t)$. The Cauchy-Schwarz inequality states that

$$\begin{aligned} &| \textstyle\int_{\mathbf{R}^2 \times \mathbf{R}} d^2\vec{k} \; d\omega \; \widehat{\Psi}\left[\vec{k}, \omega\right] \widehat{\tilde{s}}\left(\vec{k}, \omega\right) |^2 \\ &\leq \textstyle\int_{\mathbf{R}^2 \times \mathbf{R}} d^2\vec{k} \; d\omega \; |\widehat{\Psi}\left[\vec{k}, \omega\right]|^2 \; \int_{\mathbf{R}^2 \times \mathbf{R}} d^2\vec{k} \; d\omega \; |\widehat{\tilde{s}}\left(\vec{k}, \omega\right)|^2 \ , \end{aligned} \tag{10}$$

where $\tilde{s}(\vec{x}, t)$ is a band-limited version of $s(\vec{x}, t)$ with one or several moments equal to zero. Then, equality proceeds if $\widehat{\Psi}(\vec{k}, \omega) = c \; \widehat{\tilde{s}}\left(\vec{k}, \omega\right)$. This inequality provides some starting conditions for the wavelet transform to perform matched filtering or correlation. The analyzing wavelet has to be matched to the object with respect to its spectrum and its motion. In our case, the unique optimum to be tuned must correspond to the trajectory. This enables a stable and unambiguous tracking procedure. This important property must then be analytically demonstrated for each family of wavelets when applied to the particular motion under investigation. This equation and all its related theory remain valid in our case and interconnect our analysis problem not only to the theory developed for mechanical systems but also to optimum control. The equations and the algorithms that have been developed to recursively construct the optimum control, apply readily to this problem. Let us mention the Kalman filter and Bellman's algorithm (Viterbi algorithm).

5. DETECTION AND TRACKING

The detection of moving objects relies on extracting local maxima in the velocity representation, $E = f(\vec{v}, a)$,

$$E(\vec{v}, a) = \int_{\tau=0}^{\tau=T} \int_{\vec{b}=\vec{b}_{min}}^{\vec{b}=\vec{b}_{max}} | < \Psi_{\vec{b},\tau,\vec{v},a} | s > |^2 \; d\tau d^2\vec{b} \tag{11}$$

i.e. from the energy density computed by integrating the energy of the CWT over the space and the length of the scene. This technique effectively characterizes all the moving objects and the velocities.

The tracking strategy is based on combining *Kalman filters* and *CWTs*. The state of the Kalman filter is composed of all the wavelet parameters. Usually, Kalman filters are characterized by two equations, a *state equation* and an *observation equation*. The state equation is an adaptive predictor that updates the state $U(n)$ of the filter

$$\hat{U}(n) = \Phi(n, n-1) U(n-1) + W(n) \ , \tag{12}$$

where $\hat{U}(n)$ is the state prediction at step n and $W(n)$ the prediction error. Φ is the transition matrix or the feedback matrix of the Kalman filter. If the state is well-chosen (i.e. the CWT matches the signal), the predictor behaves as a Markov process, and the prediction error is a zero-mean Gaussian process. In the case of an analysis with Galilean wavelets, the state parameters are composed of the set $(\vec{b}, \tau, \vec{v}, a, \theta)$ and the prediction step n is the image interval. For other CWTs (like the accelerated family), the prediction step can involve several images, typically tens of them. The CWT is then used at each step n as a motion analyzer to determine the exact state values of the Kalman filter $U(n)$. A *gradient algorithm* works in the neighborhood of the predicted state $\hat{U}(n)$ to locate the exact state $U(n)$ composed of the parameters that maximize the following energy density

$$MAX \quad E(\vec{b}, \tau, \vec{v}, a, \theta) = | < \Psi_{\vec{b},\tau,\vec{v},a,\theta} | s > |^2 \ . \tag{13}$$

The observation equation also exploits the CWT as a motion-based extraction tool tuned to the current exact state parameters. The CWT captures and isolates the selected objects from the scene s to provide a display I,

$$I(n; \vec{b}, \tau) = < \Psi_{\vec{b},\tau,\vec{v}=\vec{v}_{opt},a=a_{opt},\theta=\theta_{opt}} | s > + V(n; \vec{b}, \tau) \ . \tag{14}$$

I is the segmented image of the selected object, displayed alone at its correct location; s is the original signal under analysis, and V is the noise produced by the optical sensors.

6. MORLET WAVELET AND APPLICATIONS

The applications presented in this paper for detection and tracking has been performed with the Galilean CWT. An anisotropic *Morlet wavelet* is admissible as a mother wavelet in the Galilean family; it defines a non-separable filter

$$\begin{aligned} &\Phi(\vec{x}, t) \\ &= e^{i\vec{k}_0 \vec{X}} \; e^{-\frac{1}{2} < \vec{X} \mid C\vec{X} >} - e^{-\frac{1}{2} < \vec{k}_0 \mid D\vec{k}_0 >} \; e^{-\frac{1}{2} < \vec{X} | C\vec{X} >} \end{aligned}$$

where $\vec{X} = (\vec{x}, t)^T \in \mathbf{R}^2 \times \mathbf{R}$, C is a positive definite matrix and, $D = C^{-1}$. For $2D + T$ signals, $\left[C = \begin{pmatrix} 1/\epsilon_x & 0 & 0 \\ 0 & 1/\epsilon_y & 0 \\ 0 & 0 & 1/\epsilon_t \end{pmatrix} \right]$ where the ϵ factors introduce anisotropy in the wavelet shape. Figures 1 and 2 show the energy density of the Morlet wavelet in the Fourier domain at velocity $\vec{v} = (1, 0)$. A high selectivity or anisotropy $\epsilon_t = 1000$ has been applied to flatten the wavelet along the velocity plane. Figure 4 presents the issue of the motion detection applied to the synthetic scene displayed in Figure 3. Figures 5 and 6 present the tracking of one accelerated object captured out of five others.

7. CONCLUSIONS

Several families of spatio-temporal CWTs have been proposed in this paper as tools to analyze spatio-temporal signals with respect to mechanical criteria. Among them, the *Galilean wavelet transform* is tuned to velocities and uniform translation motion. We have shown how that CWT family can handle detection and tracking applications. We believe, at this point, that the approaches based on CWTs have promise in the area of motion tracking. Tracking has also been shown possible even under severe noise conditions, and even when occlusions occur.

REFERENCES

[1.] M. Duval-Destin and R. Murenzi "Spatio-Temporal Wavelet: Application to the Analysis of Moving Patterns", *in Progress in Wavelets Analysis and Applications (Proc. Toulouse 1992) Y. Meyer and S. Roques, Editors, Ed. Frontières, Gif-sur-Yvette, pp. 399-408, 1993.*

[2.] S.T. Ali, J.-P. Antoine, J.-P. Gazeau "Square integrability of group representations on homogeneous spaces. I. Reproducing triples and frames. II. Coherent and quasi-coherent states. The case of the Poincaré group", *Ann. Inst. H. Poincaré, Vol. 55, pp. 829-855 & 857-890, 1991.*

[3.] D. Martin "Manifold Theory, An introduction for Mathematical Physicists", *Ellis Horwood, England, 1991.*

[4.] L. G. Weiss "Wavelets and Wideband Correlation Processing", *IEEE Signal Processing Magazine, pp. 13-32, January 1994.*

[5.] J.-P. Leduc and C. Labit "Spatio-Temporal Wavelet Transforms for Image Sequence Analysis", VIII European Signal Processing Conference, EUSIPCO-96, Trieste, Italy, 10-13 September 1996, 4 pp.

[6.] J.-P. Leduc "Discrete and Continuous Spatio-Temporal Wavelet transforms", admitted for publication in IEEE Transactions on Signal Processing.

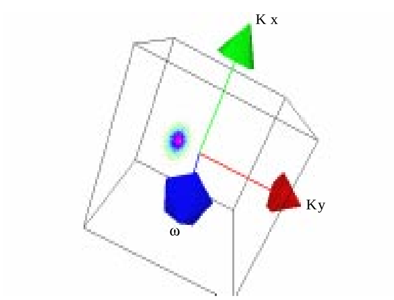

Figure 1. Galilean wavelet in velocity plane (1,0).

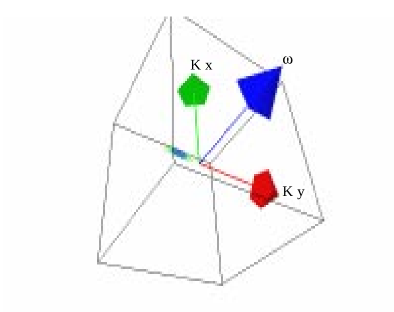

Figure 2. Galilean wavelet in velocity plane (1,0).

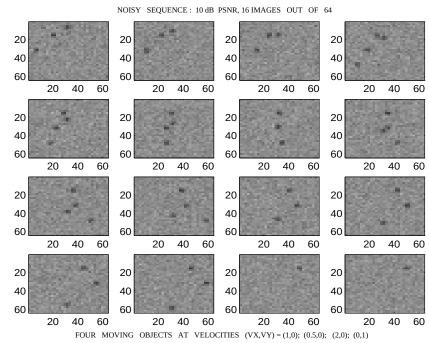

Figure 3. synthetic noisy image sequence.

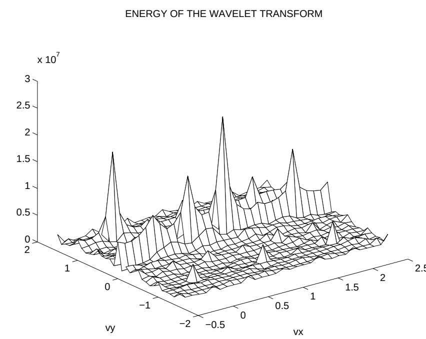

Figure 4. Velocity detection in the noisy sequence: $\vec{v} = (v_x, v_y) = (0, .5)$, $(0, 1)$ (0.2), $(1, 0)$.

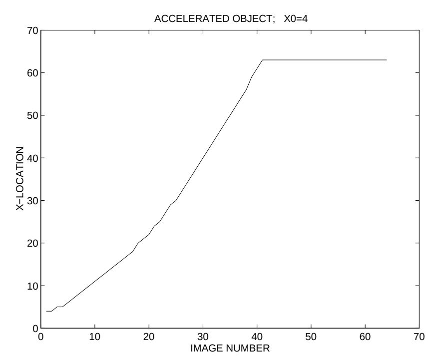

Figure 5. Selective trajectory construction (remark: the upper bound image is located at $x = 64$).

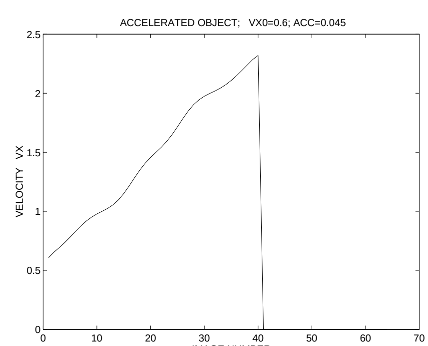

Figure 6. Selective velocity tracking.