Spatio-Temporal Continuous Wavelet Transforms for Motion-Based Segmentation in Real Image Sequences. *

Mingqi Kong †, Jean-Pierre Leduc *, Bijoy K. Ghosh †, Victor M. Wickerhauser *
Washington University in Saint Louis
† Department of Systems Science and Mathematics
* Department of Mathematics
One Brookings Drive, Campus Box 1040
Saint Louis, MO 63130

Abstract

The purpose of this paper is to develop a motion-based segmentation for digital image sequences that is based on continuous wavelet transform. Continuous wavelet transform allows estimating the motion parameters on all the moving discontinuities, edges and boundaries in the image sequence. The important fact in our case is that this technique provides all the information of motion parameter estimates and edge locations at once without going back and forth refining the segmentation and the motion parameter estimation. Also, this is achieved without involving any point/block corresponding techniques in our algorithm. The edges and the motion parameter estimates are calculated locally on small windows or pixels in the image planes by maximizing the square of the modulus of the wavelet transform. A clustering procedure allows separating all the detected edges into clusters of homogeneous motion. Building a ridge-skeleton on the reconstructed edges in each cluster provides the ultimate motion-based segments or partition. The algorithm was simulated using real traffic image sequences acquired by a mobile camera and proved to be accurate and robust.

1 Introduction

In this paper, we present a motion-based segmentation of image sequences [1] using spatio-temporal continuous wavelet transforms [11]. The goal is to partition the image sequence into segments that have different motion characteristics and properties. It is assumed that the digital signals of interest are acquired from a moving camera or a planar sensor and structured as digital image sequences. The motion studied in the signal are two-dimensional spatial projections in time of motion taking place in a three-dimensional space. For example in Figures 1 and 4, the three-dimensional translational motion of the cars on the horizontal plane of the road is transformed in the image sequence into an translational and deformational motion. In this case, the deformational component is an expansion related to the velocity component orthogonal to the camera.

The spatio-temporal continuous wavelets that are used in this paper are the Galilean wavelets that have been described in[11]. These wavelets behave as matched filters and perform minimum-mean-squared-error estimations of velocity, orientation, scale, spatio-temporal positions [11]. They actually act upon the signal as a probe (a spatio-temporal band-pass filter) that estimates simultaneously the accurate edge locations, the orientation and the velocity up to the related uncertainties of all the moving discontinuities in the scene. Therefore, this method is not subject to the classic chicken-and-egg problem described in [3] where the algorithm proceeds back and forth in between the motion estimation and the motion-based segmentation. Indeed, in our case, both information of motion and edges are computed from the wavelet transform simultaneously. The edge location, velocity and orientation correspond to the maxima of the squared modulus of the continuous wavelet transforms. A clustering process is then necessary to perform the region extraction of all the moving features as maximum regions satisfying motion-related homogeneity. In this paper, we extend the application of the continuous wavelet transforms to complex motions simultaneously involving translational and deformational motion and moving camera. As consequence, the velocities are significantly changing with time and with

*This material is based upon work supported in part by DOE under grant # $DE-FG02-90ER14140$

0-8186-8821-1/98 $10.00

the position along the feature boundary. Nevertheless, the Galilean wavelet remains quite accurate in the determination of all the velocities and edge locations. It turns out from this work that the Galilean wavelet is a very efficient tool.

The approach of motion filtering based on continuous wavelet transforms considered in this paper differs fundamentally from other techniques that have been proposed in the literature such as those based on gradient-based optical flow, block matching, pel-recursive, Bayesian model and Markov random field (MRF) models[2], [4], [5], [6], [7]. The continuous wavelet transform provides motion estimations that are robust not only against image noise and blur but also against motion noise (i.e. jitter) [11]. Moreover, as a result of both the spatio-temporal filtering and the interpolation wavelet properties, the wavelet technique can resolve temporary occlusion problems. The continuous wavelet transform presented in this paper are squared integrable Lie group representations of the Galilei group.

Eventually, simulation results are presented on real image sequences involving traffic analysis observed by a moving camera (Figures 1 and 4). Further work to be presented deals with the analytical calculations to sketch the sensitivity, the selectivity and the accuracy of the motion estimation. The uncertainties embedded in the signal between translation, velocity and orientation are given in [10], [12].

2 General Algorithmic Aspects

The algorithm is based on three steps without feedback. It allows us to derive a motion-based segmentation in any image plane of the scene. The first step consists in applying the continuous wavelet transform to estimate the velocity and the edges in small neighborhoods. For that purpose, the image is partitioned into grids of small square areas on which edges are detected and reconstructed and on which velocity and orientations are simultaneously estimated. The practical size of these square areas (blocks or windows) may vary from four by four, two by two or even one single pixel. At this stage, we get the entire information in the image plane concerning all the edge locations with an estimate of orientation and velocity.

The second step consists in clustering and partitioning the images in terms of their motion content. This part of the algorithm extracts the moving features out of the background. The segmentation of interest in our case is not object-based but instead motion-based to insulate features of coherent motion. This further allows classifying the scene in terms of its motion content. Hierarchical clustering may be achieved when a rank is stated on the motion interest. The clustering thresholds are the only parameters that need to be adjusted in this algorithm. Alternatively, the number of clusters can also be imposed. The clustering defines a partition of windows on the image plane where the moving features have to be located.

The third step consists in locating the moving features within the clusters by building connected ridge skeletons in each clustered windows. Indeed, all the edge locations correspond to the locus of maxima of the square modulus of the wavelet transform tuned to the optimum motion parameters.

3 Spatio-temporal Continuous Wavelet Transform

The Galilean wavelet referential transformation is given by

$$\begin{pmatrix} \vec{x}' \\ t' \\ 1 \end{pmatrix} = \begin{pmatrix} aR(\theta) & \vec{v} & \vec{b} \\ 0 & 1 & \tau \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \vec{x} \\ t \\ 1 \end{pmatrix} \quad (1)$$

where the parameters of interest in our case are the spatial translation $\vec{b} \in \mathbf{R}^2$, the temporal translation $\tau \in \mathbf{R}$, the velocity $\vec{v} \in \mathbf{R}^2$, the scale $a \in \mathbf{R}_+ \backslash \{0\}$, the orientation $\theta \in [0, 2\pi)$.

Let $g = \{\vec{b}, \tau, \vec{v}, a, \theta\}$ be a group element. The group representation in the spatio-temporal Hilbert space is:

$$[T_g \hat{\psi}](\vec{k}, \omega) = a^{-1/2} e^{i(\vec{k}\cdot\vec{b}+\omega\tau)} \hat{\psi}(aR(\theta)\vec{k}, \vec{v}\cdot\vec{k}+\omega) \quad (2)$$

where the hat $\hat{}$ stands for Fourier transform, $\vec{k}$ and ω stand as the spatial and the temporal frequencies, the wavelet Ψ is a *mother wavelet*, it must satisfy the condition of admissibility calculated from square integrability [12].

The applications presented in this paper are based on Morlet wavelets. An anisotropic *Morlet wavelet* is admissible as a continuous wavelet in the rotational and translational family. The still 2D+T Morlet wavelet defines a non-separable filter

$$\begin{aligned} \hat{\Psi}(\vec{k}, \omega) &= \hat{\Psi}(\vec{K}) \\ &= |det(D)|^{\frac{1}{2}} \left(e^{-\frac{1}{2}(\vec{K}-\vec{K}_0)^T D(\vec{K}-\vec{K}_0)} \right. \\ &\quad \left. - e^{-\frac{1}{2}\vec{K}_0^T D \vec{K}_0} e^{-\frac{1}{2}\vec{K}^T D \vec{K}} \right) \end{aligned} \quad (3)$$

where $\vec{K} = (\vec{k}, \omega)^T \in \mathbf{R}^2 \times \mathbf{R}$, $\vec{K}_0 = (k_{0x}, k_{0y}, \omega_0)^T$, D is a positive definite matrix. For $2D + T$ signals, $\left[D = \begin{pmatrix} \epsilon_x & 0 & 0 \\ 0 & \epsilon_y & 0 \\ 0 & 0 & \epsilon_t \end{pmatrix} \right]$ where the ϵ factors introduce anisotropy in the wavelet shape.

The Galilean wavelet transform $W[s;g]$ of the signal $s(\vec{x},t)$, with group element g is defined as an inner product that is computed in the Fourier domain. It is defined in [11] as

$$\begin{aligned} W[s;g] &= c_{\Psi}^{-1/2} < \hat{\Psi}_{\vec{b},\tau,\vec{v},a,\theta} \mid \hat{s} > \\ &= c_{\Psi}^{-1/2} \int_{\mathbf{R}^2 \times \mathbf{R}} d^2\vec{k} d\omega [T_g \hat{\psi}] \left(\vec{k},\omega\right) \bar{\hat{s}} \left(\vec{k},\omega\right) \end{aligned} \quad (4)$$

where the overbar $\bar{}$ stands for the complex conjugate.

4 Motion Estimation and Image Reconstruction

The image plane is divided into a grid of small windows or square blocks whose size is four by four, two by two or one pixel according to the requested accuracy. Let $\mathcal{B}_{ij}$ be the (i,j)th block on the nth image plane $\tau = t_n$ at a given scale $a = a_n$, the estimated velocity and orientation are obtained as

$$(\vec{v}^*, \theta^*) = \arg \max_{\vec{v},\theta} \sum_{\vec{b} \in \mathcal{B}_{ij}} | < \hat{\Psi}_{\vec{b},\tau=t_n,\vec{v},a=a_n,\theta} | \hat{s} > |^2 \quad (5)$$

i.e. by summing the square modulus of the wavelet transform (computed on the whole sequence) on the small windows $\mathcal{B}_{ij}$ of interest at image $\tau = t_n$.

To reconstruct the edges in the windows $\mathcal{B}_{ij}$ of interest at image $\tau = t_n$, the intensities $I(\vec{b})$ of the wavelet transform of the optimum velocity and orientation are stored inside the windows $\mathcal{B}_{ij}$, this means

$$I\left(\vec{b}, \tau = t_n\right) = | < \hat{\Psi}_{\vec{b},\tau=t_n,\vec{v}=\vec{v}^*,a=a_n,\theta=\theta^*} | \hat{s} > |^2, \quad \forall \vec{b} \in \mathcal{B}_{ij} \quad (6)$$

Performing the reconstruction on all the blocks of the image leads to a reconstructed image. Examples are provided in Figures 2 and 5.

5 Motion-based Clustering

In this step, we cluster the space of the optimum parameters derived from the previous section. This space is made of the optimum velocities and the coordinates of the blocks in the image plane. Scale is fixed and orientation is optimized with velocity. A preliminary procedure consists in removing all the blocks with small reconstructed image intensities below a threshold. These blocks do not contain any significant edges and any useful velocity information. Consequently, our clustering consists in partitioning the four-dimensional space of $\vec{b} = (x,y), \vec{v}^* = (v_x^*, v_y^*)$. This procedure is performed in each image plane i.e. at any time $\tau = t_n$ on the scene. According to the Huyghen's principle of decomposing the variance, we look for the minimum set of clusters that maximizes the internal cluster homogeneity and the external cluster heterogeneity.

The partition obtained by the clustering locates the areas of the features moving with similar motion properties as shown in Figure 3.

6 Motion-based Segmentation

The motion-based segmentation operates on the areas defined by the clustering procedure to extract location of the moving feature contours. In each cluster, the contours of the moving feature correspond to the ridges of intensity defined on the reconstructed images obtained in Section (4) Equation (6). The edges obtained from the reconstructed image are blurred as a result of the band-pass filtering effect of the wavelet[8],[9]. To detect the exact location of those edges, we introduce a sliding window technique. The sliding direction is that computed from the optimum local orientation, θ^*, during the wavelet analysis and we record the maximum value of the intensity inside the sliding window. This leads to isolating the different segments in motion. The exact boundaries of two moving features have been processed in Figure 6. Iterating the procedure of the motion-based segmentation on each image plane of the scene allows us to build the motion tube corresponding to each car separately. The tube is then a segment of coherent motion that spans over the scene and lasts until the feature leaves the conic field of visibility of the sensor plane. Building these segments allows performing further motion-adapted signal processing in the tube: temporal interpolation, de-noising, predicting or coding.

7 Analysis and Calculation

Even in case of signals with complex motion, velocity $\vec{v}$ is still constant within small neigborhoods of time, motion is locally translational in the first order of the Taylor expansion. Let Δ be a short period around time τ, we show that the Fourier transform is still centered on the exact velocity plane. This shows the robustness of the Galilean wavelet in performing motion estimation at any τ. The accuracy of the estimation is related to the sampling density. Let

$$\begin{aligned} x' &= x - v_x(t-\tau) \\ y' &= y - v_y(t-\tau) \\ t' &= t - \tau \end{aligned}$$

Let $s(\vec{x},t)$ be the 2D+T still signal, and $\tilde{s}(\vec{x},t)$ be its moving version. Then the Fourier $\hat{\tilde{s}}(\vec{k},\omega)$

$$= \int\int\int_{-\infty}^{+\infty} dxdydt\ \tilde{s}(\vec{x},t)\ e^{-i(k_x x + k_y y + t\omega)}$$

$$= \int\int\int_{-\infty}^{+\infty} dxdydt\ s(x - v_x(t-\tau), y - v_y(t-\tau)) \left[u\left(t - (\tau - \tfrac{\Delta}{2})\right) - u\left(t - (\tau + \tfrac{\Delta}{2})\right)\right] e^{-i(k_x x + k_y y + t\omega)}$$

$$= \int\int\int_{-\infty}^{+\infty} dx'dy'dt' \; s(x',y')[u(t'+\tfrac{\Delta}{2}) - u(t'-\tfrac{\Delta}{2})] \; e^{-i(k_x(x'+v_xt')+k_y(y'+v_yt')+\omega(t'+\tau))}$$

$$= e^{-i\omega\tau} \int\int_{-\infty}^{+\infty} dx'dy' \; s(x',y') \; e^{-i(k_xx'+k_yy')} \; [\int_{-\infty}^{+\infty} dt' \; [u(t'+\tfrac{\Delta}{2}) - u(t'-\tfrac{\Delta}{2})] \; e^{-i[(v_xk_x+v_yk_y+\omega)t']}]$$

$$= \hat{s}(k_x,k_y) \cdot \Delta \cdot Sa(\tfrac{\Delta\,\omega'}{2}) \; e^{-i\omega\tau} \qquad (7)$$

where $\omega' = v_xk_x + v_yk_y + \omega$, $Sa(x)$ is a sinc function defined by $Sa(x) = \frac{\sin(x)}{x}$. Hence, the Fourier transform is centered still around $\omega' = 0$, which corresponds to the velocity plane. When $\Delta \to 0, \Delta \cdot Sa(\frac{\Delta\,\omega'}{2}) \to \delta(\omega')$.

If we assume that both camera and tracked object are moving at constant velocity on a horizontal plane, then the component of the relative velocity orthogonal to the sensor plane transforms in the image sequence into an initial scale and expansion as $v_z \to a(t) = a_0(1+\sum_k s_kt^k)$ and the two other components of the relative velocity $\vec{v} = (v_x, v_y)$ are captured in the scene up to a scaling factor. The sensor field of visibility is a cone. Basic calculations from projective geometry show that the scale at time $t = 0$ is $a_0 = \frac{w}{S_0}$ i.e. the ratio of the visible width of the rigid object and the diameter of the visibility cone at the initial location. At time $t = t_n$, the scale is $a_n = a_0\frac{1}{1-(\frac{v_z}{d}t_n)} = a_0(1 + \frac{v_z}{d}t_n + (\frac{v_z}{d}t_n)^2 + ...)$. The signal captured by the camera from the rigid object in motion has the form $s(\vec{x}\,', t')$ with $\vec{x}\,' = a_0(1+\sum_k s_kt^k)\vec{x} - \vec{v}t - \vec{b}$ and $t' = t - \tau$. The Galilean wavelet matches to (and estimates) the first order of the temporal Taylor expansion i.e. to the apparent edge velocity $\vec{v}^* = a_0(s_1 + \sum_{k=1}^{\infty}(k+1)s_{k+1}t^k)\vec{x} + \vec{v}$ $= a_0(\frac{v_z}{d} + \sum_{k=1}^{\infty}(k+1)(\frac{v_z}{d})^{k+1}t^k)\vec{x} + \vec{v}$ a function of position and time that corresponds to the experimental observations. The same reasoning applies for objects moving with any translational and/or rotational motion involving constant or accelerated components of motion. Similarly, accelerated wavelet transforms defined in [10] will match to (and estimate) the apparent accelerations corresponding to the higher orders of temporal Taylor expansion.

8 Conclusions

In this paper, we are developing a new method to achieve motion-based segmentation of real images. The method is original in the sense that it is based on a spatio-temporal continuous wavelets that provide velocity, location and orientation of all the discontinuities embedded in the digital scene simultaneously. The technique has been shown to be robust against image noise, motion jitter and temporary occlusion [12]. Further studies will develop projective continuous wavelet transforms that optimally estimate, track, segment and reconstruct moving object according to their actual motion parameters in 3-D+T made of translational and rotational components.

References

[1] A. Murat Tekalp "Digital Video Processing", *Prentice-Hall, 1995.*

[2] J. L. Barron, D. J. Fleet, and S.S. Beauchemin "Systems and experiment: Performance of optical flow techniques ", *International Journal of Computer Vision, vol. 12, no. 1, pp.43-77, 1994.*

[3] A. Mitichie "Computational Analysis of Visual Motion", *Plenum, New York, 1994.*

[4] J.-M. Odobez and P. Bouthemy "MRF-based Motion Segmentation exploiting a 2D motion model robust estimation", *Proceedings, ICIP'95, p.3 vol(xliii+664+666+672), 628-31, vol3, Oct., 1995.*

[5] N. J. Konard and E. Dubois "Bayesian estimation of Motion vector fields ", *IEEE Transaction on Pattern Analysis and Machine Intelligentce. vol. 14, no. 7, pp. 910-927, sep. 1992.*

[6] J. N. Driessen, L. Boroczky, and J. Biemond "Pelrecursive motion field estimation from image sequences ", *Journal of Visual Communication and Image Reproduction. vol. 2 no. 3, pp. 259-280, May. 1991.*

[7] B. Liu and A. Zaccarin "New fast algorithms for the estimation of block motion vectors ", *IEEE Transactions on Circuits and Systems for Video Technology, vol. 3 no. 2, pp. 148-157, Apr. 1993.*

[8] P. Guillemain and R. Kronland-Martinet "Ridges associated to continuous linear time-frequency representations of asymptotic and transient signals", *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, pp. 177-80, Paris, June 1996.*

[9] Salari, E. and Siy, P. "The Ridge-Seeking Method for Obtaining the Skeleton of Digital Image", *IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-14, No. 3, pp. 524-528, 1984.*

[10] J.-P. Leduc, J. Corbett, M. Kong, V. Wickerhauser, B.K. Ghosh "Accelerated Spatio-Temporal Wavelet Transforms: An iterative trajectory estimation", *Proceedings of ICASSP-98, Seattle, Vol. 5, pp. 2781-2784, May 1998.*

[11] J.-P. Leduc "Spatio-Temporal Wavelet Transforms for Digital Signal Analysis", *Signal Processing, Elsevier, Vol. 60 (1), pp. 23-41, July 1997.*

[12] M. Kong, J.-P. Leduc, B. K. Ghosh, J. Corbett and V. Wickerhauser "Wavelet Based Analysis of Rotational Motion in Digital Image", *Proceedings of ICASSP-98, Seattle, Vol. 5, pp. 2777-2780, May, 1998.*

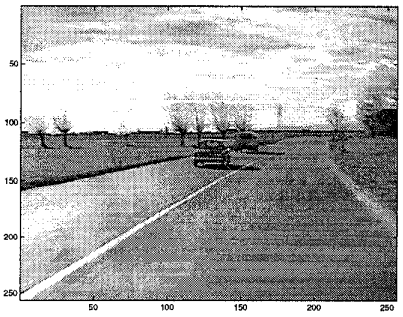

Figure 1: The 15th image of the traffic image sequences.

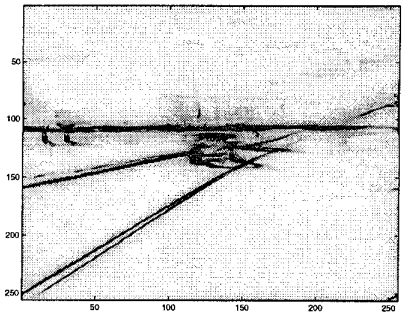

Figure 2: The reconstructed 15th image for the sequences according to Equation (6).

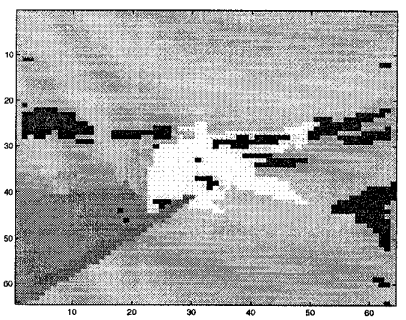

Figure 3: The clustering of the 20th image according to velocity and position.

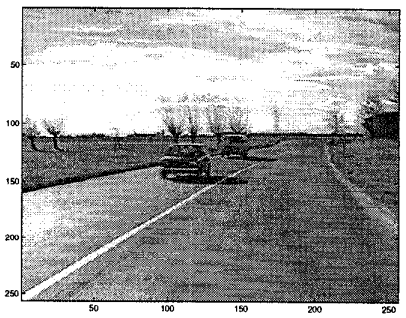

Figure 4: The 20th image of the traffic image sequences.

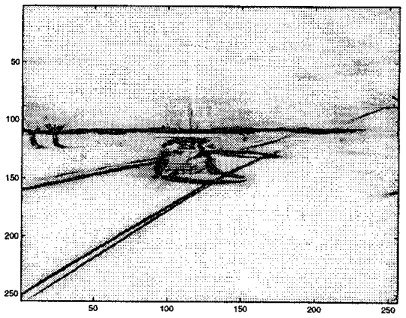

Figure 5: The reconstructed 20th image for the sequences according to Equation (6).

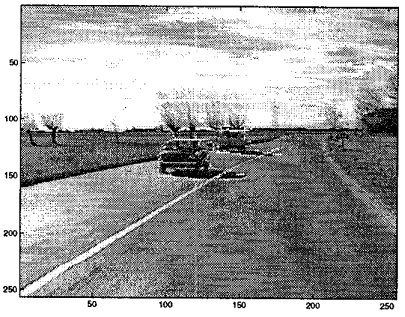

Figure 6: The exact boundaries for both moving cars in the 20th image.