MOTION-COMPENSATED ADAPTIVE WAVELET FILTERING FOR IMAGE SEQUENCE PROCESSING. *

J.-P. Leduc J.-M. Odobez C. Labit
IRISA / Centre INRIA - Rennes
Campus de Beaulieu
avenue du Général Leclerc, F-35042 Rennes, France
Tel:+(33)-99-847425 Fax:+(33)-99-847171
Email: < name > @irisa.fr

ABSTRACT

This paper intends to present new approaches in the field of motion-compensated spatio-temporal filters applied to digital image sequences. In time-varying imagery, the temporal correlation information of pixel intensities is folded by motions which may originate from both camera and object displacements. Motion-compensated filters are here defined as temporal filters applied along assumed motion trajectories. As a matter of fact, this paper deals with three-dimensional spatio-temporal filters and aims at generalizing the motion-compensated temporal filtering process as the product of two distinct operators. The first operator depends only on the estimated motion parameters derived from both motion-based image segmentations and parametric affine modelings of regions in motion. The second operator analyzes only the correlations of image-by-image intensities measured along the assumed motion trajectories. Multiresolution filters or wavelets may be consequently applied along the motion trajectories to produce optimum and adaptive resulting procedures for purposes like spatio-temporal prediction, interpolation and smoothing. In this paper, applications are provided to cover the field of image sequence coding and interpolation.

Key Words: parametric models, motion estimation, spatio-temporal filtering, multiresolution, adaptive filtering.

1. INTRODUCTION

This paper is concerned with the three-dimensional or spatio-temporal filtering of digital image sequences. Whereas our research work stems originally from coding algorithms designed for bit-rate compressions of moving picture sequences, the topic stands as a key point to several applications involving *prediction*, *interpolation* for either coding or format conversions and *filtering* for either signal analysis in coding or noise reduction of noisy signals (restoration and reconstruction).

Digital image sequences are three-dimensional signals composed of a succession of scenes. Each scene is made of several consecutive images which can be segmented into a set of three-dimensional areas characterized by specific motion models significantly different from one area to the other. Motion is represented by time-varying parameters. The motion models considered in this paper belong to the family of the affine models. Even if a three-dimensional rigid motion gives at least rise to a quadratic model in the image plane, affine flow models should recover the essential part of the motion. Exploiting a motion-based segmentation algorithm similar to [8], images are first partitioned into different temporally linked moving regions. Efficient robust algorithms [10] allow us to estimate the motion parameters within each region. Eventually, this motion characterization defines a dense motion vector field and consequently one trajectory for each pixel located in a given region. Arising from motions and also from possible variations of source intensity, the illumination of each point can vary in time along its trajectory. These *a priori* unknown variations of the illuminating intensity leads to probabilistic models expressing the correlation between point intensity in consecutive image planes. In order to perform appropriate temporal interpolation, prediction or filtering, optimum convolutional operators may be adapted to the local correlative structure of the signal along the motion trajectories. In real-time moving picture sequences, the intensity correlation is naturally folded in the signal according to the local motion. The underlying idea exploited in this paper is therefore to extract the motion from the image sequence to catch the effective intensity correlation before filtering. Images are initially segmented into regions on which affine motion parameters are estimated between each pair of images to allow the trajectory determinations. Therefore, before temporally interpolating, filtering or predicting a point intensity of concern by means of a convolutional operator (analyzing operator), the trajectory of that point is first computed time-backward or/and time-forward to derive the coordinates of the point in the image planes of interest. Since those coordinates do not correspond to the sampling lattice, the intensity of the point at each image is computed by a two-dimensional spatial interpolation (unfolding operator).

Earlier works in the field of three-dimensional subband filtering have been reported successively in [1] to [3]. In [4] and [5], attempts to combine three-dimensional subband filters and motion compensation using block matching algorithms have brought successful results in increasing the bit-rate compression and producing simultaneously more pleasing images. Kronander in [2] and Dubois in [6] have eventually introduced the *concept of motion-compensated filtering* as a way of using the motion estimates not to make predictions but to filter the pixel values along their motion trajectory. This contrasts to the earlier research works where a motion-compensated prediction was build and encoded on the basis

*The following text presents a research work supported by the Commissions of the European Communities-DG XII under a fellowship ERB4001GT930967 of the Human Capital and Mobility.

of motion estimates. Other recent research works [7] are also worth mentioning; they concern directional wavelets in N-dimensional spaces and also spatio-temporal wavelets as tools for motion tracking applications. The actual originality brought in the research work presented in this paper stays in combining new results obtained in motion estimation with affine models to those of adaptive wavelet filters in order to derive adaptive motion-compensated filter banks. This study leads to discriminate in the signal the motion features from the correlation characteristics and, consequently, to apply distinct estimators and operators in the signal analysis. Motion parameters are exploited to unfold the signal from motion and the correlation parameters to decompose the signal with optimally adapted wavelets.

2. MOTION-BASED IMAGE SEGMENTATIONS AND PARAMETRIC MODELS OF MOTION

Two-dimensional polynomial motion models and, more specifically the affine flow representations, are considered in this work to characterize image-to-image motions. Displacement vectors are given by

$$D\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} t_x \\ t_y \end{pmatrix} + \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x - x_g \\ y - y_g \end{pmatrix} \quad (1)$$

where $D()$, (x, y), (x_g, y_g) are respectively the displacement vector, the coordinates of the displaced point and the coordinates of the center of gravity for the region to which point (x, y) belongs. t_x and t_y are the translation components and a, b, c and the other affine motion parameters. This linear affine model stands experimentally as a good trade-off between complexity and representativeness [8, 9]. It can take into account different motion components, namely the translation, the rotation, the scaling and the shear (linear deformation). Usually two versions of the affine model are taken into consideration, a simplified model with four parameters t_x, t_y, r and θ defined as $a = r\ cos\theta$, $c = r\ sin\theta$, $b = -r\ sin\theta$, $d = r\ cos\theta$ (translation, rotation and scaling) and the complete model composed of six parameters (two rotations and scalings to incorporate the shear).

The goal of the motion-based analysis is to estimate a map of regions $\{R_k\}$ with 2D affine motion descriptors $\{\Theta_k\}$ which best model the motion activity between pairs of images. This structure $\{R_k, \Theta_k\}$ provides a compact representation for trajectory determination. In our case, the segmentation algorithm proceeds in three basic steps. The first one consists in estimating the motion model on each region in the motion-oriented projection of the segmentation map at the previous instant. Owing to robust estimation [10], the inaccuracy of this map as well as local violation of the constraint of brightness constancy along motion trajectory is not critical. The segmentation map is obtained using a statistical regularization based on multiscale Markov Random Fields, which uses motion observations as well as their reliability. The third step consists in the detection of the areas where the motion description is not correct [10], and in creating new regions if necessary. In the analys process, the segmentation is computed at the resolution of pixel accuracy.

3. ADAPTIVE FILTERING

Motion-compensated adaptive filtering as studied in this paper results (Figure 2) from the product of two operators, namely a *two-dimensional spatial interpolation* in the image planes and a *temporal wavelet filtering*. The first operator only depends on the *affine parameters of motion* (six or four parameters) estimated either time-forward or time-backward according to the application in vigor (e.g. the causality required in image coding imposes a time-backward estimation, the temporal interpolation applications use both). Eventually, the set of all affine motion descriptors of the regions defines a dense motion vector field. Owing to these motion vector fields, spatio-temporal motion trajectories of the image points may be traced out through the sequence to determine in each image plane their exact coordinates in each image. Point intensity at each coordinate location in image planes are determined by a two-dimensional spatial interpolation based on cubic convolution and interpolation kernels (Figure 1). Let us here remark that the geometrical operator can be represented in whole generality by either linear or non-linear geometrical spatial transformations of image planes. Indeed, if f(x,y) and g(u,v) are respectively the original and the transform image coordinates, then

$$[x\ y\ 1] = [u\ v\ 1] \begin{pmatrix} \frac{dx}{du} & \frac{dy}{du} & 0 \\ \frac{dx}{dv} & \frac{dy}{dv} & 0 \\ x_0 & y_0 & 1 \end{pmatrix} \quad (2)$$

Linear transform is characterized by constant matrix elements to represent translation, rotation, scaling and shear (six parameters or degree of freedom maximum). In the case of non-linear transformations, the matrix elements are no longer constant. Non-linear distortions are considered in general up to deformations of the second order derivative of the Taylor expansion.

The second operator is uniquely characterized by the *variations of illumination* of the point of concern along its own trajectory. It refers to a temporal filter applied along the estimated motion trajectory. In whole generality, that filter can also be either linear or non-linear. An application of image sequence coding will be presented below in the paper. In interpolation applications, the Galerkin method of approximation can be used as described in [11]. Owing to both properties of orthogonality and of multiresolution, wavelet functions adequately fit as trial or basis functions for the Galerkin approximation method (equivalent to a least-squares method). The intensity function along the motion trajectory ϕ is here approximated by $\hat{\phi}$. If any function f_0 can be found which takes on known points the same values of ϕ and if a set of M orthogonal basis functions $\{\Psi_m\ ;\ m = 1, ..., M\}$ is introduced such that $\Psi_m = 0\ ;\ m = 1, .., M$ on known points of ϕ, then at any point to be interpolated, we can approximate to ϕ by $\phi = f_0 + \sum_{m=1}^{M} a_m \Psi_m$. In Galerkin method [11], the unknown coefficients $a_m\ ;\ m = 1, ..., M$ are solved with a symmetric matrix equation with noticeable computation advantages.

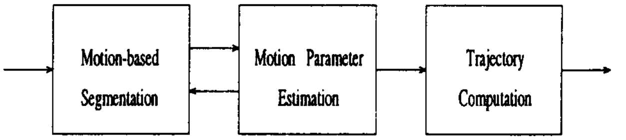

Figure 1. Motion identification, estimation and trajectory computation.

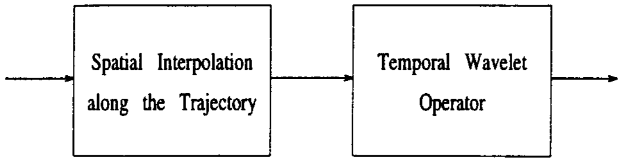

Figure 2. Motion-compensated filtering.

3.1. Motion-compensated Filtering

The motion compensation process implemented by filtering along the trajectories is studied as the product of two operators. Without loss of generality, the scene can be considered as made of only one object and then the first operator can be considered as a transformation T_s (unfolding) of the sequence $f(x,y,t)$ such that $T_s: \ f(x,y,t) \rightarrow f[T_s(x,y,t),t]$. Wavelet transform is further applied along the time axis. An important question can be risen about the preservation of the wavelet properties through an affine transformation. We except the trivial case of the translation and focuse on the four matrix coefficients of 1. If the affine matrix can be decomposed into elementary unitary transformations such as rotation, dilatation, spatio-temporal deformation at constant volume, reflection, the resulting operator is still a wavelet. Moreover, though the spatio-temporal filters are built in separable way (2D+T), the resulting filter is locally separable for still objects and becomes non-separable when the object is put in motion.

3.2. Adaptive Wavelet Filtering along the Motion Trajectories

Any kind of three-dimensional (2D+T) spatio-temporal filter can be taken here into consideration (separable, non-separable, FIR or IIR, orthogonal or bi-orthogonal wavelets,...) to yield any filtering performance or criterion. Since the motion effects have been removed by unfolding the three-dimensional image signal, spatial and temporal correlations may be similarly dealt with.

To illustrate this issue, a temporal multiresolution decomposition for image sequence coding will be considered with the design of a two-channel bi-orthogonal FIR filter bank which yields perfect reconstruction and linear-phase conditions. For obvious reasons, perfect reconstruction is required in the image transmissions; linear-phase responses guarantee a pure constant delay. At least one degree of regularity is required to design wavelets, this implies for the high-pass filter response to satisfy the constraint of having a zero-mean response, that is $\sum_n (-1)^n h_1(n) = 0$. $h_0(n)$ and $h_1(n)$ stand respectively as the low-pass and high-pass analysis impulse responses of length n_0 and n_1. The tilded versions $\tilde{h}_0(n)$ and $\tilde{h}_1(n)$ correspond to the synthesis filter bank. Choosing pure-delay and linear-phase filter banks, the design is constrained to $\tilde{h}_0(n) = (-1)^n h_1(n)$ and $\tilde{h}_1(n) = -(-1)^n h_0(n)$. Two types of nontrivial filter banks have been characterized which lead to linear phase filters. Type A refers to odd-length and symmetric filter responses and type B to even-length and antisymmetric filter responses. Moreover, the filter bank can be adapted to the input signal statistics in order to concentrate a maximum of energy on the low-pass output response (decorrelation). This is achieved with the best approximation when imposing simultaneously a *quasi-orthogonality* (i.e. when $h_0(n)$ is as close as possible to $\tilde{h}_0(n)$) and a ***minimum energy at the high-pass filter output***, that is

$$MIN\left[H_1^T \ R_{XX} \ H_1 \ + \ \| H_0 - \tilde{H}_0 \|^2\right] \qquad (3)$$

where H_1 and H_0 are the impulse response vectors and R_{XX} the auto-covariance matrix of the input signal. Perfect reconstruction imposes the general constraint equations of the form $A \ H_1 = m$ with $A =$

$$\begin{vmatrix} h_0(1) - h_0(0) & 0 & \cdots & 0 \\ h_0(3) - h_0(2) & H_0(1) - h_0(0) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \ldots & \ldots & h_0(l_0-2) - h_0(l_0-1) & h_0(l_0) \end{vmatrix}$$

and $m = [0 \ \ldots \ 1]$. $l_0 = \frac{N_0}{2} - 1$ and $l_1 = \frac{N_1}{2} - 1$ for type B solutions. The practical way to solve that problem with quadratic cost function and non-linear constraints requires an iterative procedure described in [12] under the Lagrange-Newton algorithm. This leads to a classical Lagrangian function which can be written as $L(H_0, H_1, \lambda) = \Phi(H_0, H_1) - \lambda^T(AH_1 - m)$. The iterative method towards optimum is analytically described in [12]. This approach generally leads to better filtering performances. These aspects are fully developed in paper [13].

4. APPLICATIONS - RESULTS

An application in image coding (for bit-rate compression) is presented as a matter of example, Figures 3 and 4 present both low-pass and high-pass subbands of a temporal dyadic decomposition for image analysis and coding without motion-compensated filtering and Figures 5 and 6 the corresponding results with motion-compensated filters. Such decompositions can be iterated in a whole multiresolution application. It turns out that motion-compensated filtering concentrates as expected more energy into the low-temporal subband (better decorrelation). For example, comparative results obtained by applying four coding schemes are encouraging. The four coding schemes are as follows: (1.) 2D spatial coder, (2.) 3D coder, (3.) 3D coder with motion-compensated filters, (4.) adaptive version of case 3. The test sequences were *"interview"* and *"caltrain"* and the PSNR was about 34 dB for each results. The mean bit rate are for cases 1. to 4.s in bit/pel: 0.61, 0.47, 0.40, 0.36 for *"caltrain"* and 0.59, 0.48, 0.41, 0.38 for *"interview"*.

5. CONCLUSIONS

This paper has addressed a general problem of image sequence filtering which finds in fact applications in many imaging problems like coding, restoration, enhancement and format conversion since it can deal with prediction and interpolation. Intrinsically, moving image sequences have to cope simultaneously with objects in motion and with their variations of illumination. The problem of motion-compensated filtering has been here generalized by splitting the filtering process into two distinct operators. The first deals only with motion and the second deals only with the intensity correlation along the trajectory to derive the filtered output. The solution of this problem has combined results from different theories; these are namely the affine parametric modeling, the operator theory, the adaptive and optimum filtering theory, the approximation, the multiresolution and the wavelet theory. Examples in the field of three-dimensional image sequence coding and of spatio-temporal interpolation applications have been provided. Since optimum smoothing along trajectories are nothing else than an application of the Wiener-Kalman filtering theory, other multiresolution applications can also be easily derived from this paper.

Figure 3. Low-frequency temporal subband (not motion-compensated).

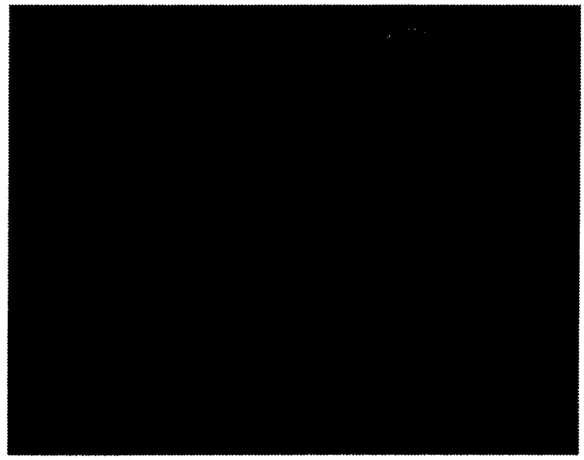

Figure 4. High-frequency temporal subband (not motion-compensated).

Figure 5. Low-frequency temporal motion-compensated subband.

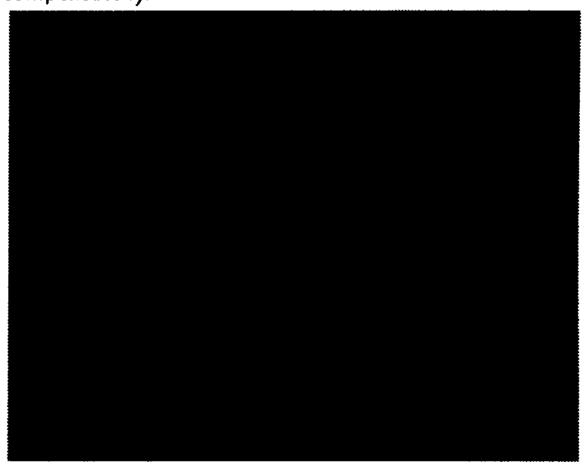

Figure 6. High-frequency temporal motion-compensated subband.

REFERENCES

[1] Karlsson and Vetterli "Three-Dimensional Subband Coding of Video." *Proceedings of ICASSP-88, 1988.*

[2] T. Kronander "Some aspects of Perceptual Based Image Coding", *Ph.D. Thesis, Linkoping University, Sweden, 1989.*

[3] A. S. Lewis and G. Knowles "Video Compression Using 3-D Wavelet Transforms." *Electronic Letters, March 1990, pp. 396-398.*

[4] J.-R. Ohm "Temporal Domain Subband Video coding with Motion Compensation." *Proceedings of IEEE ICASSP-92, Vol. III, pp. 229-232, 1992.*

[5] J. W. Woods and G. Lilienfield "Experiments in Motion Compensated 3-D Subband Coding." *Proceedings of the Eighth Workshop on IMDSP, session C2, pp. 84-85, 1993.*

[6] E. Dubois "Motion-Compensated Filtering of Time-Varying Image." *Multidimensional Systems and Signal Processing, 3, pp. 211-239, 1992.*

[7] J.-P. Antoine, R. Murenzi "Directional wavelets in higher dimensions." *Clark Atlanta University, Center for Theoretical Studies of Physical Systems, technical report, P3-94, 1994.*

[8] P. Bouthemy and E. François "Motion segmentation and qualitative dynamic scene analysis from an image sequence." *International Journal of Computer Vision, Vol. 10, No. 2, pp. 157-182, April 1993.*

[9] H. Nicolas and C. Labit "Global motion identification for image sequence analysis and coding." *Proceedings of ICASSP-91, Toronto, Vol. IV, pp. 2825-2828, May 1991.*

[10] J.-M. Odobez and P. Bouthemy "Detection of Multiple Objects using Multiscale MRF with Camera Motion Compensation." *Proceedings of ICIP-94, Austin, Texas, Vo II, pp. 257-261, November 1994.*

[11] O. C. Zienkiewicz and K. Morgan "Finite Elements and Approximation." *John Wiley & Sons, New York, 1983, Chapter 2, pp. 38-94.*

[12] R. Fletcher "Practical Methods of Optimization." *John Wiley & Sons, New York, 1981, Vol. 2.*

[13] J.-P. Leduc, J.-M. Odobez and C. Labit "Adaptive Three-dimensional motion-compensated wavelets for Image Sequence Coding." *Submitted to Signal Processing: Image Communication, December 1994.*