International Journal of Computer Vision 46(2), 129–155, 2002

Hierarchical Estimation and Segmentation of Dense Motion Fields

ETIENNE MÉMIN
IRISA/Univ. Rennes I, Campus de Beaulieu, 35042 Rennes Cedex, France
memin@irisa.fr

PATRICK PÉREZ
Microsoft Research, St George House 1, Guildhall Street Cambridge, CB2 3NH, UK
pperez@microsoft.com

Received July 26, 2001; Revised May 22, 2001; Accepted July 30, 2001

Abstract. In this paper we present a comprehensive energy-based framework for the estimation and the segmentation of the apparent motion in image sequences. The robust cost functions and the associated hierarchical minimization techniques that we propose mix efficiently non-parametric (dense) representations, local interacting parametric representations, and global non-interacting parametric representations related to a partition into regions. Experimental comparisons, both on synthetic and real images, demonstrate the merit of the approach on different types of photometric and kinematic contents ranging from moving rigid objects to moving fluids.

Keywords: apparent motion, robust discontinuity-preserving estimation, motion-based segmentation, hierarchical non-linear minimization, dense and parametric representations

1. Introduction

Among early vision problems, the estimation and the segmentation of the apparent motion from an image sequence is particularly intricate. It is a two-fold problem which lies at the heart of most tasks of video analysis. It is thus a critical part of a number of computer vision applications such as motion detection in a scene, 3D motion and scene structure recovery, obstacle avoidance in robotics, etc. (see for example, Mitiche and Bouthemy (1996) for a review on motion analysis issues).

Stemming either from a discrete Markovian framework or from a deterministic continuous one, energy-based models are very appealing to handle in a versatile way high-dimensional inverse problems. For motion analysis purposes, such models have been thoroughly investigated. They usually rely on the brightness constancy assumption (i.e., the image irradiance of a physical point does not change within a certain time interval) combined with some *a priori* knowledge on the displacement field. This prior is either captured *locally* by a smoothness term in the cost function (Horn and Schunck, 1981), or defined more *globally* as a parametric representation of the unknown motion (Adiv, 1985; Ayer and Sawhney, 1995; Bergen et al., 1992; Bouthemy and Francois, 1993). These two types of prior have their own advantages and drawbacks. Contrary to the local smoothing approach, the parametric modelization relies on large spatial supports of estimation. This type of modelization is thus usually dedicated to motion-based segmentation where areas with kinematic meanings have to be extracted from the images (Ayer and Sawhney, 1995; Bergen et al., 1992; Bouthemy and Francois, 1993; Murray and Buxton, 1987), provided that such regions with consistent motions exist in the sequence at hand. This type of approach is also expected to be more reliable, provided that the selected parameterization makes sense from a physical point of view. Local non-parametric models are, in that sense, more versatile since they only

capture smoothness assumptions on the desire solution. Besides, they are independent of any partition of the image, and they can easily handle local features of the motion field such as discontinuities (Black and Rangarajan, 1996). For these reasons, they are usually involved in *dense* motion estimation (Black and Rangarajan, 1996; Cohen and Herlin, 1999; Kornprobst et al., 1999; Mémin and Pérez, 1998a).

The limitations of each approach can be illustrated in the case of images involving fluids. In sequences of this nature, like those encountered in meteorology (atmospheric satellite images) or fluid mechanics (videos of wind tunnel or water tank experiments), it is common to observe very low photometric contrasts. Although the brightness constancy assumption can be turned into a more appropriate "transport" model (Corpetti et al., 2000; Schunck, 1986) in this specific context, the data model will always be plagued by the absence of consistent photometric information to rely on. In that case, the parametric approach, with its use of extended estimation supports, could seem more appropriate. However, the physics of the scene makes its use delicate: in fluid motions there is no real objects or motion regions with borders, and the involved motions can be much more complex than those captured by standard parametric models.

Based on these preliminary remarks, we aim at mixing a local non-parametric smoothing and a more global parametric representation. We actually present two different (and not exclusive) methods to reach that goal. The first one concerns a particular constrained minimization technique used with an energy-based dense motion estimation model. The second method deals with an energy-based model for the joint estimation-segmentation of the apparent motion. In both approaches, whose cost functions are partly the same, robust penalty functions are used to deal with the various deviations from the selected models. We also propose to build on the so-called "semi-quadratic" rewriting of such robust functions with auxiliary weights, by using the auxiliary weights as a device to couple the different variables of the problem.

As already mentioned, energetic formulations can be viewed either from a continuous angle or from a discrete one. The former kind of approaches implies continuous functionals, variational calculus, deterministic partial differential equations, and discretization schemes (finite differences, finite elements), whereas the latter type of formalism is often related to Markov random fields and Bayesian inference. The two viewpoints provide different insights into a given problem, as well as different mathematical tools to cope with the various issues at hand. Despite their differences, it is known that these two types of approaches often lead to very similar discrete implementations. Although our setting relies more on a discrete philosophy, we shall discuss the connection between the two points of view, in terms of both models and algorithms.

The paper is divided into three main parts. In Section 2 we focus on a robust energy-based model for the incremental dense estimation of the apparent motion field with preservation of its discontinuities. To cope with the associated minimization we introduce an efficient tailor-made hierarchical technique which combines different and varying parameterizations of the unknown field. The compromise between local dense methods and global parametric approaches is thus introduced via the minimization process. In Section 3 we show how the former energy function can be extended to estimate at the same time a motion-based segmentation of the scene. The resulting joint estimation-segmentation model introduces another mix between local smoothness and region-wise parameterization. The global minimization is performed with a natural extension of the hierarchical optimization technique developed in the previous part. The last part (Section 4) is devoted to experimental results. The two approaches are validated qualitatively and quantitatively on real world sequences of quite different natures. Systematic comparisons are also provided on a synthetic benchmark, including an assessment of the sensitivity of the proposed techniques to the value of the different parameters.

2. Robust Estimation of Dense Motion

The dense estimation of the apparent motion aims at recovering a displacement field $\boldsymbol{w} = \{\boldsymbol{w}_s, s \in S\}$ over the rectangular pixel lattice S, based on the luminance function $f(t) = \{f(s,t), s \in S\}$ at two consecutive instants t and $t+1$. Assuming the temporal constancy of the brightness for a physical point between the two images, one gets:

$$\forall s \in S,\ f(s + \boldsymbol{w}_s, t+1) = f(s,t), \qquad (1)$$

which is highly nonlinear w.r.t. the unknown displacement vector $\boldsymbol{w}_s$ at location s. A first-order expansion of the left-hand-side provides the standard optic flow

constraint equation (OFCE):

$$\nabla f(s,t+1)^T \boldsymbol{w}_s + f(s,t+1) - f(s,t) = 0, \quad (2)$$

where ∇f stands for the spatial gradient of f. For this approximation to be valid, the unknown displacement $\boldsymbol{w}_s$ should remain in the "domain of linearity" of the luminance function at location s. This is particularly unlikely to hold around sharp edges (where large gradients imply reduced linearity domains), and for large displacements. These limitations are usually circumvented by conducting an incremental estimation through a multiresolution hierarchy of sequences (Black and Anandan, 1996; Enkelmann, 1988). We, too, stick to that multiresolution setup involving a pyramidal decomposition of the images. Even if we do not make it explicit, we shall assume throughout to be working at a given resolution of such a multiresolution structure. One has to keep in mind that all definitions and derivations are thus meant to be reproduced at each resolution level according to a coarse-to-fine strategy.

2.1. *Incremental Energy-Based Model*

The incremental estimation assumes that a preliminary estimate $\boldsymbol{w} = \{\boldsymbol{w}_s, s \in S\}$ of the unknown vector field is available (e.g., from an estimation at lower resolution or at previous instant). A refinement is sought in terms of an increment field $\mathsf{d}\boldsymbol{w} \in \Omega \subset (\mathbb{R} \times \mathbb{R})^S$. Based on the linearization of the constancy brightness assumption (1) with respect to that increment, instead of the total field as in (2), a robust energy-based refinement can be defined as:

$$\widehat{\mathsf{d}\boldsymbol{w}} = \arg\min_{\mathsf{d}\boldsymbol{w}\in\Omega} [H_1(\mathsf{d}\boldsymbol{w}) + H_2(\mathsf{d}\boldsymbol{w})], \quad (3)$$

with (Black and Anandan, 1996; Mémin and Pérez, 1998a):

$$H_1(\mathsf{d}\boldsymbol{w}) \triangleq \sum_{s\in S} \phi_1([\nabla f(s+\boldsymbol{w}_s, t+1)^T \mathsf{d}\boldsymbol{w}_s + f_t(s,\boldsymbol{w}_s)]^2), \quad (4)$$

$$H_2(\mathsf{d}\boldsymbol{w}) \triangleq \alpha \sum_{\langle s,r\rangle\in\mathcal{C}} \phi_2(\|(\boldsymbol{w}_s + \mathsf{d}\boldsymbol{w}_s) - (\boldsymbol{w}_r + \mathsf{d}\boldsymbol{w}_r)\|^2). \quad (5)$$

The first term constitutes the *data-model* which rules the dependency between the unknown displacement field and the data. The second term captures a smoothness prior on the total displacement. It is made up of a sum of local discrepancy penalties taken over the set $\mathcal{C}$ of all the pairs of neighboring pixels (for a chosen neighborhood system, which is often of first or second order). The parameter $\alpha > 0$ balances the contributions of the two terms. In the data-model $f_t(s,\boldsymbol{w}_s) \triangleq f(s+\boldsymbol{w}_s, t+1) - f(s,t)$ denotes the displaced frame difference, and ϕ_1 and ϕ_2 are two increasing concave functions that soften the plain quadratic penalties such as to limit the influence of large residuals. These two functions make the model robust to large deviations from the first-order smoothness assumption and from the brightness constancy assumption respectively.

A so-called semi-quadratic formulation of penalties of the form $\phi(.^2)$, with ϕ concave, can be obtained (Black and Rangarajan, 1996; Charbonnier et al., 1997; Geman and Reynolds, 1992): if $\lim_{0^+}\phi' < \infty$, and $\lim_{+\infty}\phi' = 0$, there exists an increasing function ψ such that $\phi(x^2) = \min_{z\in(0,1]}[\tau z x^2 + \psi(z)]$, where $\tau \triangleq \lim_{v\to 0^+}\phi'(v)$, i.e, $\phi(.^2)$ is the inferior envelope of a family of parabolas continuously indexed by an *auxiliary variable* (or weight) z lying in $(0,1]$.[1] The minimizer is given by $\arg\min_{z\in(0,1]}[\tau z x^2 + \psi(z)] = \frac{1}{\tau}\phi'(x^2)$.

Using this reformulation result, the minimization of $H_1 + H_2$ can be replaced by the minimization in $(\mathsf{d}\boldsymbol{w}, \delta, \beta)$ of an augmented cost function $\mathcal{H} \triangleq \mathcal{H}_1 + \mathcal{H}_2$ with:

$$\mathcal{H}_1(\mathsf{d}\boldsymbol{w},\delta) = \sum_{s\in S}[\tau_1\delta_s[\nabla f(s+\boldsymbol{w}_s,t+1)^T\mathsf{d}\boldsymbol{w}_s + f_t(s,\boldsymbol{w}_s)]^2 + \psi_1(\delta_s)], \quad (6)$$

$$\mathcal{H}_2(\mathsf{d}\boldsymbol{w},\beta) = \alpha \sum_{\langle s,r\rangle\in\mathcal{C}} [\tau_2\beta_{sr}\|(\boldsymbol{w}_s + \mathsf{d}\boldsymbol{w}_s) - (w_r + \mathsf{d}\boldsymbol{w}_r)\|^2 + \psi_2(\beta_{sr})], \quad (7)$$

where $\tau_1 \triangleq \lim_{0^+}\phi_1'$, $\tau_2 \triangleq \lim_{0^+}\phi_2'$, and $\delta = \{\delta_s, s\in S\}$ and $\beta = \{\beta_{sr}, \langle s,r\rangle \in \mathcal{C}\}$ are two sets of auxiliary variables lying within $(0,1]$ and respectively attached to the pixel grid and to the edge grid. This new minimization can then be led alternatively with respect to $\mathsf{d}\boldsymbol{w}$ and to the weights: the energy $\mathcal{H}$ is quadratic in $\mathsf{d}\boldsymbol{w}$ and the corresponding minimization amounts to a standard weighted least squares problem; conversely, $\mathsf{d}\boldsymbol{w}$ being frozen, the best weights are obtained in closed form (Mémin and Pérez, 1998a). The convergence of this alternate scheme is guaranteed, to a global minimum if $\phi(.^2)$ is convex (Charbonnier et al., 1997), and to a local minimum otherwise (Delanay and Bresler, 1998).

In practice, the displaced frame $f(s+\boldsymbol{w}_s, t+1)$ is computed through a backward registration of $f(s, t+1)$ associated with a bilinear interpolation. The spatial gradients are then computed on this image with the derivative filters proposed in Viéville and Faugeras (1992).

This robust energy-based modeling provides a generic dense estimator which can be applied to image sequences of various natures provided that they contain sufficient photometric contrast or texture. Wherever this condition is not met within large areas, the robust smoothness term might not be strong enough to propagate the estimates obtained at the border of these regions toward inner locations where data cannot be exploited. However, the experiments reported in Section 4 demonstrate that this problem is effectively tackled by the hierarchical piece-wise parametric minimization we now introduce. Other gains will be obtained in terms of global quality of estimates and computational load.

2.2. *Piece-Wise Parametric Constraint*

Let us assume that the pixel grid is divided into a collection of patches. Let $\mathcal{B} \triangleq \{\mathcal{B}_n,\ n = 1, \ldots, N\}$ be this partition and E the edge set of the associated connectivity graph.[2] We consider piece-wise parametric increment fields for this partition. They are defined as:

$$\forall n = 1 \ldots N,\ \forall s \in \mathcal{B}_n, \quad \mathsf{d}\boldsymbol{w}_s = \Phi_n(\boldsymbol{\theta}_n, s), \tag{8}$$

where $\boldsymbol{\theta}_n$ is a p_n-dimensional parameter vector and the Φ_n's are interpolation functions which can be different from one patch to another. The whole increment field can then be expressed $\mathsf{d}\boldsymbol{w} = \Phi(\boldsymbol{\theta})$ with $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_1^T \cdots \boldsymbol{\theta}_N^T)$ lying in parameter space Γ. The full-rank function Φ is the interpolator between the reduced subspace Γ and the original configuration space Ω. It is a one-to-one mapping from Γ into the constrained configuration subset $\mathrm{Im}\Phi \subset \Omega$.

The constrained minimization of $\mathcal{H}$ in $\mathrm{Im}\Phi$ is equivalent to a new minimization defined on Γ:

$$\min_{\mathsf{d}\boldsymbol{w} \in \mathrm{Im}\Phi} \mathcal{H}(\mathsf{d}\boldsymbol{w}, \delta, \beta) = \min_{\boldsymbol{\theta} \in \Gamma} \underbrace{\mathcal{H}(\Phi(\boldsymbol{\theta}), \delta, \beta)}_{\triangleq \mathcal{H}^*(\boldsymbol{\theta}, \delta, \beta)}. \tag{9}$$

The new energy function $\mathcal{H}^*$ is readily derived from the original one (6–7). Denoting $\mathcal{C}_n \triangleq \{\langle s, r\rangle \in \mathcal{C} : \langle s, r\rangle \subset \mathcal{B}_n\}$ the set of the neighboring pixel pairs included in patch $\mathcal{B}_n$, and $\mathcal{C}_{nm} \triangleq \{\langle s, r\rangle \in \mathcal{C} : s \in \mathcal{B}_n, r \in \mathcal{B}_m\}$ the set of neighboring site pairs straddling adjacent patches $\mathcal{B}_n$ and $\mathcal{B}_m$ (see Fig. 1), one can show that this new

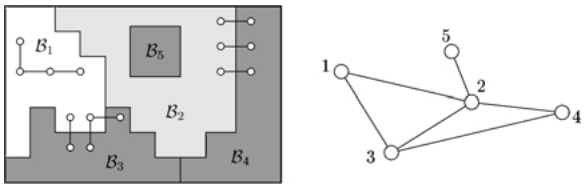


Figure 1. Image partitioning and associated notations: Example of image partition $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3, \mathcal{B}_4, \mathcal{B}_5\}$ (with examples of neighboring pixel pairs belonging to $\mathcal{C}_1$, $\mathcal{C}_{13}$, and $\mathcal{C}_{24}$ respectively), and associated adjacency graph with edge set $E = \{(1, 2), (1, 3), (2, 3), (2, 4), (2, 5), (4, 3)\}$.

energy is similarly composed of two terms, $\mathcal{H}^* = \mathcal{H}_1^* + \mathcal{H}_2^*$, defined as:

$$\begin{aligned}
\mathcal{H}_1^*(\boldsymbol{\theta}, \delta) &= \sum_{n=1}^{N} \sum_{s \in \mathcal{B}_n} \tau_1 \delta_s [\nabla f(s + \boldsymbol{w}_s, t+1)^T \\
&\quad \times \Phi_n(\boldsymbol{\theta}_n, s) + f_t(s, \boldsymbol{w}_s)]^2 + \psi_1(\delta_s), \\
\mathcal{H}_2^*(\boldsymbol{\theta}, \beta) &= \alpha \Bigg[\sum_{\langle n,m\rangle \in E} \sum_{\langle s,r\rangle \in \mathcal{C}_{nm}} \tau_2 \beta_{sr} \| (\boldsymbol{w}_s + \Phi_n(\boldsymbol{\theta}_n, s)) \\
&\quad - (\boldsymbol{w}_r + \Phi_m(\boldsymbol{\theta}_m, r)) \|^2 + \psi_2(\beta_{sr}) \\
&\quad + \sum_{n=1}^{N} \sum_{\langle s,r\rangle \in \mathcal{C}_n} \tau_2 \beta_{sr} \ \| (\boldsymbol{w}_s + \Phi_n(\boldsymbol{\theta}_n, s)) \\
&\quad - (\boldsymbol{w}_r + \Phi_n(\boldsymbol{\theta}_n, r)) \|^2 + \psi_2(\beta_{sr}) \Bigg].
\end{aligned} \tag{10}$$

Note that the first term in the definition of $\mathcal{H}_2^*$ is reminiscent of the "skin and bones" model introduced by Ju et al. (1996).

Minimizing $\mathcal{H}^*$ provides a piece-wise parametric increment field where different parameterizations are combined. Note that, contrary to what is done in parametric *segmentation* approaches based on independent region-wise parametric models, the different parameter vectors $\boldsymbol{\theta}_n$'s interact here through the smoothness term that enforces continuity at patch frontiers. In addition, contrary to splines based methods (Musse et al., 1999; Szeliski and Shum, 1996) for which parameters also interact, the presence of a robust smoothness term allows here to introduce spatial discontinuities between patches via the variables β_{sr}'s lying along patches' frontiers.

In the next section we show how this constrained optimization can be easily embedded in a hierarchical optimization framework.

2.3. *Hierarchical Constrained Optimization*

We now consider a sequence of partitions. The constrained optimization previously described can be successively used with each of these partitions, thus providing a hierarchical optimization scheme: the original optimization problem of $\mathcal{H}$ is replaced by a succession of constrained minimizations. Let $\{\mathcal{B}^\ell, \ell = L \ldots 0\}$ be the family of partitions.[3] Let E^ℓ and S^ℓ, $\ell = L \ldots 0$, be respectively the edge set and the N_ℓ-vertex set of the associated connectivity graph. The interpolation functions Φ^ℓ associated to the sequence of partitions are chosen such that the size of $\Omega^\ell \triangleq \mathrm{Im}\Phi^\ell$ decreases as ℓ increases.[4]

The constrained optimization in Ω^ℓ is equivalent to the minimization of the new energy function:

$$\mathcal{H}^\ell(\boldsymbol{\theta}^\ell, \delta, \beta) \triangleq \mathcal{H}(\Phi^\ell(\boldsymbol{\theta}^\ell), \delta, \beta), \tag{11}$$

defined, as concerns the unknown increment field, over a reduced parameter space Γ^ℓ, whereas the unknown weights δ and β, as well as the data f, and the field $\boldsymbol{w}$ to be refined remain the same (i.e., defined on the original grid S). A diagrammatic view of this setting is given by:

$$\begin{array}{ccc} \Gamma^\ell & \xrightarrow{\Phi^\ell} & \Omega^\ell \\ & \searrow^{\mathcal{H}^\ell} & \downarrow \mathcal{H} \\ & & \mathbb{R}^+ \end{array} \tag{12}$$

Based on this family of energy functions $\mathcal{H}^\ell$, we now define our minimization scheme as a recursive sequence (from $\ell = L$ to $\ell = 0$) of optimization problems of reduced complexity:

$$(\hat{\boldsymbol{\theta}}^\ell, \hat{\delta}, \hat{\beta}) = \arg\min_{\boldsymbol{\theta}^\ell, \delta, \beta} \mathcal{H}^\ell(\boldsymbol{\theta}^\ell, \delta, \beta),\ \ell = L \ldots 0, \tag{13}$$

where the field to be refined at level ℓ, $\boldsymbol{w}^\ell \triangleq \boldsymbol{w}^{\ell+1} + \Phi^{\ell+1}(\hat{\boldsymbol{\theta}}^{\ell+1})$, is deduced from the estimate at level $\ell + 1$, and the initial field $\boldsymbol{w}^L$ comes from an estimation at a coarser resolution or from a given initialization.

Each of these successive minimizations is processed in terms of iteratively reweighted least squares initialized by $\boldsymbol{\theta}^\ell \equiv \mathbf{0}$ and provides the increment field $\Phi^\ell(\hat{\boldsymbol{\theta}}^\ell)$. The procedure is repeated until the finest level $\ell = 0$ is reached, and the motion field finally recovered is $\boldsymbol{w}^L + \sum_{\ell=L}^{0} \Phi^\ell(\hat{\boldsymbol{\theta}}^\ell)$, which is not piece-wise parametric in general. This incremental minimization procedure can be viewed as a *hierarchical Gauss-Newton* minimization of $\sum_s \phi_1([f(s + \boldsymbol{w}_s, t + 1) - f(s, t)]^2) + \alpha \sum_{\langle s,r \rangle} \phi_2(\|\boldsymbol{w}_s - \boldsymbol{w}_r\|^2)$.[5]

Gauss-Newton techniques are second-order minimization techniques that resort to an approximation of the Hessian, which does not rely on the current residuals, as opposed to exact Newton technique. As a result, the speed of convergence is high when the residuals are small, but the process can be painfully slow, if not divergent, when the residuals are large (Thisted, 1988). Nevertheless, as it is shown in experimental Section 4, this latter behavior has not be observed for the particular Gauss–Newton minimization we use. We believe this is mainly due to the hierarchical setting which helps keeping the residuals small by an efficient guidance of the minimization process at the coarse levels.

2.4. *Linear Parameterizations and Energy Minimization*

So far, we let the nature of the parameterizations unspecified. In practice, the interpolation functions are chosen linear. The constrained increment field $\mathsf{d}\boldsymbol{w}^\ell = \Phi^\ell(\boldsymbol{\theta}^\ell)$ thus obeys

$$\forall n \in S^\ell,\ \forall s \in \mathcal{B}_n^\ell, \mathsf{d}\boldsymbol{w}_s = P_n(s)\boldsymbol{\theta}_n^\ell, \tag{14}$$

where $P_n(s)$ is 2 by p_n matrix. The corresponding parameter spaces are $\Gamma^\ell = \times_{n=0}^{N_\ell} \mathbb{R}^{p_n}$. Standard parametric models used in motion analysis correspond to $p_n = 2$, 4, 6 or 8 [1, 5]. In this work we will consider two possible parameterizations: the *constant model* (2 parameters of translation and $P_n(s) = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$) and the *affine model* (6 parameters and $P_n(s) = \left[\begin{smallmatrix} 1 & x_s & y_s & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_s & y_s \end{smallmatrix}\right]$, where x_s and y_s stand for the coordinates of pixel s). As reported in Section 4, we have investigated different combinations of these two parameterizations. A simplified affine modeling with four parameters has also been investigated in a previous work (Mémin and Pérez, 1998b), where it appeared as less interesting than the two models used here. Also, as compared to more complex models, such as the eight or twelve-parameter quadratic models, the six-parameter affine model is known to provide the best compromise between computational efficiency, robustness, and versatility (Bouthemy and Francois, 1993).

Introducing the linear parameterizations (14) within the energies (10) yields

$$\begin{aligned}\mathcal{H}_1^\ell(\boldsymbol{\theta}^\ell,\delta) &= \sum_{n\in S^\ell}\sum_{s\in\mathcal{B}_n^\ell}\tau_1\delta_s\big[\boldsymbol{\nabla} f\big(s+\boldsymbol{w}_s^\ell,t+1\big)^T \\ &\quad\times P_n(s)\boldsymbol{\theta}_n^\ell + f_t\big(s,\boldsymbol{w}_s^\ell\big)\big]^2+\psi_1(\delta_s),\\ \mathcal{H}_2^\ell(\boldsymbol{\theta}^\ell,\beta) &= \alpha\Bigg[\sum_{\langle n,m\rangle\in E^\ell}\sum_{\langle s,r\rangle\in\mathcal{C}_{nm}^\ell}\tau_2\beta_{sr}\big\|\big(\boldsymbol{w}_s^\ell+P_n(s)\boldsymbol{\theta}_n^\ell\big)\\ &\quad-\big(\boldsymbol{w}_r^\ell+P_m(r)\boldsymbol{\theta}_m^\ell\big)\big\|^2+\psi_2(\beta_{sr})\\ &\quad+\sum_{n\in S^\ell}\sum_{\langle s,r\rangle\in\mathcal{C}_n^\ell}\tau_2\beta_{sr}\big\|\big(\boldsymbol{w}_s^\ell+P_n(s)\boldsymbol{\theta}_n^\ell\big)\\ &\quad-\big(\boldsymbol{w}_r^\ell+P_n(r)\boldsymbol{\theta}_n^\ell\big)\big\|^2+\psi_2(\beta_{sr})\Bigg]. \end{aligned}\tag{15}$$

The iteratively reweighted least squares minimization applied to this energy function amounts to alternate updates of the weights and of the parameter vectors. The current parameter estimate $\boldsymbol{\theta}^\ell$ being fixed, we know that the optimal weight values are directly accessible. These values are:

$$\begin{aligned}&\forall n\in S^\ell,\forall s\in\mathcal{B}_n^\ell,\ \delta_s\\ &\quad=\frac{1}{\tau_1}\phi_1'\big[\big(\boldsymbol{\nabla} f\big(s+\boldsymbol{w}_s^\ell,t+1\big)^T P_n(s)\boldsymbol{\theta}_n^\ell\\ &\qquad+f_t\big(s,\boldsymbol{w}_s^\ell\big)\big)^2\big],\end{aligned}\tag{16}$$

$$\begin{aligned}&\forall\langle n,m\rangle\in E^\ell,\forall\langle s,r\rangle\in\mathcal{C}_{nm}^\ell,\ \beta_{sr}\\ &\quad=\frac{1}{\tau_2}\phi_2'\big[\big\|\big(\boldsymbol{w}_s^\ell+P_n(s)\boldsymbol{\theta}_n^\ell\big)-\big(\boldsymbol{w}_r^\ell+P_m(r)\boldsymbol{\theta}_m^\ell\big)\big\|^2\big],\end{aligned}\tag{17}$$

$$\begin{aligned}&\forall n\in S^\ell,\forall\langle s,r\rangle\in\mathcal{C}_n^\ell,\ \beta_{sr}\\ &\quad=\frac{1}{\tau_2}\phi_2'\big[\big\|\boldsymbol{w}_s^\ell-\boldsymbol{w}_r^\ell+(P_n(s)-P_n(r))\boldsymbol{\theta}_n^\ell\big\|^2\big].\end{aligned}\tag{18}$$

It is worth noting that according to (18), the discontinuity variables β_{sr} located into patches of $\mathcal{B}^\ell$ (i.e., $\langle s,r\rangle\in\mathcal{C}_n^\ell$ for some $n\in S^\ell$) do not depend on the translational components of $\boldsymbol{\theta}^\ell$. Indeed, the unity entries associated to these components in matrices $P_n(s)$ cancel themselves out in the matrix difference $P_n(s)-P_n(r)$. In the piece-wise constant case, the discontinuity variables therefore depend only on $\boldsymbol{w}^\ell$, and can be computed right away within the first iteration at the current grid level.

As soon as the values of all weights are computed and frozen, the energy function $\mathcal{H}^\ell(\boldsymbol{\theta}^\ell,\delta,\beta)$ is quadratic with respect to $\boldsymbol{\theta}^\ell$. Its minimization is equivalent to the resolution of a linear system whose solution is searched with a block-based Gauss-Seidel solver. Each single update of this iterative process is obtained by solving a linear equation in $\boldsymbol{\theta}_n^\ell$ for the current block $\mathcal{B}_n^\ell$. This is detailed in the Appendix A for the two different parameterizations on $\mathcal{B}_n^\ell$.

Before explaining how the model can be enriched to deal with a joint segmentation process, it is worth discussing the connection between the discrete model we have presented and its possible continuous counterpart.

2.5. *Continuous Formalization*

A continuous version of the discrete energy (6–7), which was derived by Taylor expansion around the current displacement field, corresponds to the functional

$$\begin{aligned}&\iint_{\mathcal{S}}(\tau_1\delta(\boldsymbol{x})[\boldsymbol{\nabla}\tilde{f}(\boldsymbol{x})^T\mathsf{d}\boldsymbol{w}(\boldsymbol{x})+\tilde{f}_t(\boldsymbol{x})]^2+\psi_1(\delta(\boldsymbol{x})))\,d\boldsymbol{x}\\ &+\alpha\iint_{\mathcal{S}}(\tau_2\beta(\boldsymbol{x})\|\boldsymbol{\nabla}(\boldsymbol{w}(\boldsymbol{x})+\mathsf{d}\boldsymbol{w}(\boldsymbol{x}))\|^2+\psi_2(\beta(\boldsymbol{x})))\,d\boldsymbol{x},\end{aligned}\tag{19}$$

where $\boldsymbol{w}$ and $\mathsf{d}\boldsymbol{w}$ are, momentarily, two C^1-vector fields over the continuous plane domain $\mathcal{S}\subset\mathbb{R}^2$, δ and β are two scalar fields on the same domain, and $\boldsymbol{\nabla}\tilde{f}\triangleq\boldsymbol{\nabla} f(.+\boldsymbol{w},t+1)$, $\tilde{f}_t\triangleq f(.+\boldsymbol{w},t+1)-f(.,t)$. The problem of minimizing this half-quadratic functional can be addressed in terms of alternate minimization (Deriche et al., 1995). For fixed $\mathsf{d}\boldsymbol{w}$, the Euler-Lagrange equations provide the optimal expression of functions δ and β (using $\psi'(z)=-\tau z\phi'^{-1}(\tau z)$):

$$\begin{aligned}\delta&=\frac{1}{\tau_1}\phi_1'[(\boldsymbol{\nabla}\tilde{f}^T\mathsf{d}\boldsymbol{w}+\tilde{f}_t)^2],\quad\text{and}\\ \beta&=\frac{1}{\tau_2}\phi_2'[\|\boldsymbol{\nabla}(\boldsymbol{w}+\mathsf{d}\boldsymbol{w})\|^2].\end{aligned}\tag{20}$$

The natural discretization of the first equation, which consists in taking the values of δ, $\tilde{f}_t$, and $\boldsymbol{\nabla}\tilde{f}$ at pixel locations $\boldsymbol{x}=s\in S$, is readily obtained and yields exactly the same update rule as the one stemming from the minimization of the discrete energy $\mathcal{H}$ w.r.t. $\{\delta_s\}_{s\in S}$. The same discretization scheme can be adopted for the second equation (as in Nesi, 1993). However, if the gradients are approximated by finite differences on the grid, it is simpler to have the function β discretized on the same edge lattice. The corresponding discretized update is then the same as the one that minimizes $\mathcal{H}$ w.r.t. $\{\beta_{sr}\}_{\langle s,t\rangle}$.

The weight functions δ and β being fixed, one has to deal with the minimization of a quadratic functional of $\mathsf{d}\boldsymbol{w}$. This can be conducted first by writing down the Euler-Lagrange equations as a necessary condition of optimality:

$$\tau_1 \delta \nabla \tilde{f} \nabla \tilde{f}^T \mathsf{d}\boldsymbol{w} - \alpha \tau_2 \mathrm{div}[\beta \nabla(\boldsymbol{w} + \mathsf{d}\boldsymbol{w})] = -\tau_1 \delta \tilde{f}_t \nabla \tilde{f}. \quad (21)$$

If $\boldsymbol{w}$, $\mathsf{d}\boldsymbol{w}$, and δ are discretized on S, while β, $\nabla \boldsymbol{w}$, and $\nabla \mathsf{d}\boldsymbol{w}$ are discretized on the edge grid, and the divergence operator is approximated by first-order central difference on S, this partial differential equation leads to a linear system which coincides with the one to be solved for minimizing $\mathcal{H}$ in $\mathsf{d}\boldsymbol{w}$.

We see that a standard discretization based on finite difference turns the minimization of the continuous functional into the same problem as the one issuing from the minimization of the discrete energy $\mathcal{H}$. The continuous formalism, however, allows more flexibility in the choice of the discretization schemes since the discretization step is "delayed": whereas the discrete modeling sticks right away to the pixel grid discretization, the variational approaches offer other choices, especially when the finite element method is used (within Euler-Lagrange formalism (Cohen and Herlin, 1999), or apart from it (Schnorr et al., 1996)). In many cases, however, the discretization of the original continuous model is made as simple as possible w.r.t. the pixel grid, thus yielding *in fine* the same discrete problems to be solved as those associated with the minimization of discrete energies.

As concerns the constrained minimization scheme introduced in Section 2 when dealing with the discrete energy $\mathcal{H}$, it can be viewed in the continuous framework as a Galerkin technique for solving the linear system that arises from the discretization of (21). Denote $A\mathsf{d}\boldsymbol{w} = \boldsymbol{b}$ this system. Provided that the interpolator Φ (Eq. (8)) from the reduced configuration subspace Γ into the complete configuration space Ω is linear, the standard coarse-to-fine multigrid approach (Hackbusch, 1985) relies on the resolution in Γ of the so-called Galerkin system $\Phi^T A^T A \Phi \boldsymbol{\theta} = \Phi^T A^T \boldsymbol{b}$. The solution of this equation is obviously the minimizer of the quadratic energy $\|A\Phi\boldsymbol{\theta} - \boldsymbol{b}\|^2$. In other terms it corresponds to the minimizer of $\|A\mathsf{d}\boldsymbol{w} - \boldsymbol{b}\|^2$ within subspace $\mathrm{Im}\Phi$. In the case of the simple discretization scheme mentioned earlier, $\mathcal{H}$ and $\|A\mathsf{d}\boldsymbol{w} - \boldsymbol{b}\|^2$ coincide up to an additive term independent from $\mathsf{d}\boldsymbol{w}$, and solving the Galerkin system above provides the minimizer in $\boldsymbol{\theta}$ of the reduced energy $\mathcal{H}^*$.

The whole approach to dense motion estimation we have introduced in this section could thus have been equivalently formulated in a continuous fashion, as it is done in related works (Cohen and Herlin, 1999; Deriche et al., 1995; Schnorr et al., 1996). However, this does not hold for the augmented estimation-segmentation model to be presented.

3. Joint Estimation-Segmentation

In the previous section we have described a general hierarchical method to estimate dense motion fields. We shall see with experimental results that this approach provides a family of hierarchical motion estimators which give good results on sequences involving a variety of motions. Before reporting these experiments, we now introduce an extension of the model to couple the estimation process with a motion-based partition of the image.

Motion estimation and motion-based segmentation are two tightly interwoven problems: a good estimation of the motion field (or at least a sensible approximation of it) is required to obtain a good segmentation of the different apparent motions present in the scene; conversely, a good estimation of the motion field cannot be obtained without an accurate estimation of the frontiers of the different moving objects. It is therefore natural to consider the resolution of these two problems as a whole.

This has been considered in a number of different ways and within a variety of methodological frameworks. Nevertheless, two main classes of estimation-segmentation approaches can be distinguished. The first one consists in an unilateral coupling between some motion estimate (such as sparse matchings, estimate of contour motions, or dense motion estimate) and a segmentation process (Adiv, 1985; Ayer and Sawhney, 1995; Bouthemy and Francois, 1993; Huang et al., 1995; Murray and Buxton, 1987; Oisel et al., 2000). In this class of methods, the motion cues are first extracted, and then used as the data on which the segmentation is built. The second class of methods implies a real coupling between the estimation of the motion and the extraction of a motion-based partition of the image within a joint procedure. This is usually achieved using a global energy function depending on both entities. In this context, different kinds of interactions have been recently proposed. In Stiller (1997) the frontiers

of the partition capture in a binary way the discontinuities of the dense displacement field under estimation. The motion field is thus smoothed independently from one region of the current partition to another. There is no piece-wise parametric representation on which the segmentation relies. In Black and Jepson (1996) the interaction consists in a cooperation between a dense motion field and a region-wise parametric polynomial flow. As in our approach, the motion is encouraged to have some similarity with the piece-wise parametric field associated with the segmentation. Nevertheless, the partition results from an early photometric processing, and is kept fixed afterward. The coupled model we propose is similar to the one proposed in Chang et al. (1997). In Chang et al. (1997) however, the DFD-based cost term involves a plain quadratic penalization and no Gauss-Newton like incremental linearizations are considered. This results in a huge computation load and leads sometimes to bad results.

The estimation-segmentation coupling we consider here thus belongs to the class of joint approaches. We aim at building, through a global discrete energy function, a cooperative method to estimate simultaneously a dense motion field and a motion-based segmentation. The associated minimization is solved with an extension of the hierarchical optimization scheme described in Section 2.

3.1. Compound Energy

Let $\mathcal{R} \triangleq \{\mathcal{R}_1 \cdots \mathcal{R}_p\}$ be a partition of S into an *unknown* number p of connected regions. We shall call "boundary" between regions $\mathcal{R}_i$ and $\mathcal{R}_j$ the set $\partial\mathcal{R}_{ij} \triangleq \{\langle s, r\rangle \in \mathcal{C} : s \in \mathcal{R}_i, r \in \mathcal{R}_j\}$, where we remind that $\mathcal{C}$ is the set of all neighboring pixel pairs within pixel grid S. The boundary set $\partial\mathcal{R}_{ij}$ is thus the set of the neighboring pairs straddling $\mathcal{R}_i$ and $\mathcal{R}_j$; it is non empty if $\mathcal{R}_i$ and $\mathcal{R}_j$ are adjacent for the graph on S. The set $\partial\mathcal{R} \triangleq \cup_{\langle i,j\rangle}\partial\mathcal{R}_{ij}$, where $\langle i, j\rangle$ denotes all pairs of neighboring regions of segmentation $\mathcal{R}$, stands therefore for the total frontier of the segmentation map.

The extension of the energy-based motion estimation model of previous section is obtained by incorporating two terms to the global energy function $\mathcal{H}$.[6] The first one, $\mathcal{H}_3$, specifies the mode of interaction between the segmentation and the rest of the estimation model (i.e., motion fields, weights, and data). The interaction we designed involves the total motion field both at the frontiers and inside the regions: the partition will interact with the estimation process through the discontinuity weights along these frontiers, and through a parametric goodness-of-fit criterion inside each region. The second energy term, $\mathcal{H}_4$, captures the *a priori* knowledge about the segmentation configuration.

The energy component $\mathcal{H}_3$ is composed of two terms. The first term is proportional to the sum of the β_{sr}'s averages on the individual boundaries $\partial\mathcal{R}_{ij}$'s. This term favors low values (close to zero) of discontinuity weights along the borders and guides the boundaries of the segmentation toward the most significant motion discontinuities. This constitutes an extension to an arbitrary partition of the original mechanism we introduced in Mémin and Pérez (1998a) in the case of a single closed curve. The second term enforces the likeness with a parametric representation of the dense motion field inside each region, via a robust penalization of the discrepancies by a third robust function ϕ_3. Before semi-quadratic rewriting, this new compound energy term reads:

$$\underbrace{\mu_1 \sum_{\langle i,j\rangle} \frac{1}{|\partial\mathcal{R}_{ij}|} \sum_{\langle s,r\rangle\in\partial\mathcal{R}_{ij}} \beta_{sr}}_{\text{frontiers term}} + \underbrace{\mu_2 \sum_i \sum_{s\in\mathcal{R}_i} \phi_3(\|\boldsymbol{w}_s + \mathsf{d}\boldsymbol{w}_s - P(s)\boldsymbol{\varphi}_i\|^2)}_{\text{likeness term}}, \tag{22}$$

where μ_1 and μ_2 are positive parameters, $P(s)$ is the 2×6 matrix defined in Section 2.4, and $\boldsymbol{\varphi}_i$ is the six-component parameter vector of region $\mathcal{R}_i$. Using the semi-quadratic rewriting described in Section 2.1, we get:

$$\mathcal{H}_3(\mathsf{d}\boldsymbol{w}, \mathcal{R}, \boldsymbol{\varphi}, \beta, \eta) \triangleq \mu_1 \sum_{\langle i,j\rangle} \frac{1}{|\partial\mathcal{R}_{ij}|} \sum_{\langle s,r\rangle\in\partial\mathcal{R}_{ij}} \beta_{sr} + \mu_2 \sum_i \sum_{s\in\mathcal{R}_i} [\tau_3\eta_s\|\boldsymbol{w}_s + \mathsf{d}\boldsymbol{w}_s - P(s)\boldsymbol{\varphi}_i\|^2 + \psi_3(\eta_s)] \tag{23}$$

with $\tau_3 \triangleq \lim_{0^+} \phi_3'$, $\boldsymbol{\varphi} \triangleq [\boldsymbol{\varphi}_1 \cdots \boldsymbol{\varphi}_p]$, and $\eta \triangleq \{\eta_s, s \in S\}$ is the new set of auxiliary variables which will be referred to as *parametric likeness weights*.

The energy term $\mathcal{H}_4$ captures a loose geometric prior based on the Minimum Description Length (MDL) principle (Leclerc, 1989):

$$\mathcal{H}_4(\mathcal{R}) \triangleq \lambda|\partial\mathcal{R}|, \tag{24}$$

for some $\lambda > 0$. This energy term favors short and smooth segmentation frontiers.

The whole energy function $\mathcal{H} \triangleq \mathcal{H}_1 + \mathcal{H}_2 + \mathcal{H}_3 + \mathcal{H}_4$ has now to be minimized with respect to all the unknowns. A direct minimization is obviously a very intricate problem since the two main sets of unknowns $\mathsf{d}\boldsymbol{w}$ and $\mathcal{R}$ are interacting through $\mathcal{H}_3$. Nevertheless, the minimization can be efficiently conducted through an extension of the hierarchical minimization strategy used for the motion estimation problem alone.

3.2. *Hierarchical Optimization*

As in Section 2.3, a family of finer and finer partitions $\mathcal{B}^\ell$, $\ell = L \ldots 0$, is specified and the optimization problem is solved through a sequence of constrained minimizations based on these partitions. At level ℓ, the problem is the joint estimation of an increment field $\mathsf{d}\boldsymbol{w}^\ell$ which is piece-wise parametric w.r.t. $\mathcal{B}^\ell$, and a partitioning $\mathcal{R}^\ell$ of the elements of $\mathcal{B}^\ell$. Note that $\mathcal{B}^\ell$ and $\mathcal{R}^\ell$ are two partitions of completely different nature. As explained in Section 2.3, $\mathcal{B}^\ell$ is a grid whose cells shrinks as $\ell \to 0$ and which supports the estimation of the unknowns (i.e., the increment field $\mathsf{d}\boldsymbol{w}$ and the segmentation $\mathcal{R}$) within a top-down scheme. Each region of the partition $\mathcal{R}^\ell$ is composed of one or several cells of $\mathcal{B}^\ell$. This partition $\mathcal{R}^\ell$ is aimed to break up the scene into individual kinematic components. The precision of this partition increases as $\ell \to 0$. Also $\mathcal{B}^\ell$ specifies patches on which the unknown *increment* is parametric, with the parameter vectors on neighboring patches interacting through the smoothness prior, whereas the regions of $\mathcal{R}^\ell$ support *independent* parametric representations of the *whole* field. Let us also note that, even if the increment field is constrained to be piece-wise parametric on $\mathcal{B}^\ell$, the total dense motion field does not respect this constraint in general.

In the adaptive version of the dense motion estimation alone, the patches of variable sizes constituting $\mathcal{B}^\ell$ are expected to do, partly and temporary, the job of segmentation. Now that the segmentation is explicitly and properly handled by $\mathcal{R}^\ell$, an adaptive grid partitioning seems less relevant. Hence, we chose for the $\mathcal{B}^\ell$'s the simple nested family of $2^\ell \times 2^\ell$-block partitions. For the same reason, it is sufficient to use the piece-wise constancy constraint on the increment field, i.e., at level ℓ, the increment field over one patch of $\mathcal{B}^\ell_n$ is equal to a constant two-dimension vector denoted $\mathsf{d}\boldsymbol{w}^\ell_n$.

At a given grid level ℓ, the joint configuration subset is denoted $\Omega^\ell \times \Upsilon^\ell$ (see an instance of such constrained configurations in Fig. 2). As in Section 2.3, each constrained configuration of Ω^ℓ is equivalently described

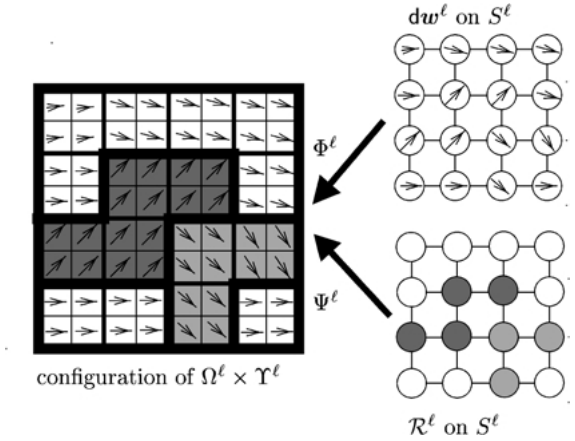


Figure 2. Multigrid setting and associated notations: (Left) example of an increment field $\mathsf{d}\boldsymbol{w}^\ell$ and a segmentation $\mathcal{R}^\ell$ in five regions, constrained to lie on a 2×2 block partition $\mathcal{B}^\ell$ of S; (Right) the associated increment field and partition on reduced grid S^ℓ.

by a reduced increment field, $\mathsf{d}\boldsymbol{w}^\ell \in \Gamma^\ell$, lying on the grid S^ℓ, with the one-to-one mapping from Γ^ℓ into Ω^ℓ being denoted Φ^ℓ. In the same way, any constrained partition of Υ^ℓ is associated with a partition into connected components of the reduced grid S^ℓ. If Λ^ℓ is the set of such partitions, the corresponding mapping from Λ^ℓ into Υ^ℓ is denoted Ψ^ℓ, with $\Upsilon^\ell = \mathrm{Im}\Psi^\ell$ (see Fig. 2).

The constrained optimization in $\Omega^\ell \times \Upsilon^\ell$ is then equivalent to the minimization of the new energy function:

$$\begin{aligned}&\mathcal{H}^\ell(\mathsf{d}\boldsymbol{w}^\ell, \mathcal{R}^\ell, \varphi, \delta, \beta, \eta)\\ &\quad\triangleq \mathcal{H}(\Phi^\ell(\mathsf{d}\boldsymbol{w}^\ell), \Psi^\ell(\mathcal{R}^\ell), \varphi, \delta, \beta, \eta),\end{aligned} \tag{25}$$

defined over $\Gamma^\ell \times \Lambda^\ell$, whereas the auxiliary variables, the data, and the field to be refined are still defined on the original grid S. The new diagram for the joint setting is:

$$\begin{array}{ccc}\Gamma^\ell \times \Lambda^\ell & \xrightarrow{(\Phi^\ell, \Psi^\ell)} & \Omega^\ell \times \Upsilon^\ell \\ & \searrow^{\mathcal{H}^\ell} & \downarrow \mathcal{H} \\ & & \mathbb{R}^+\end{array} \tag{26}$$

We now deal with the cascade of optimization problems of reduced complexity:

$$\min_{\mathsf{d}\boldsymbol{w}^\ell, \mathcal{R}^\ell, \varphi, \delta, \beta, \eta} \mathcal{H}^\ell(\mathsf{d}\boldsymbol{w}^\ell, \mathcal{R}^\ell, \varphi, \delta, \beta, \eta), \ \ell = L \ldots 0. \tag{27}$$

In the same way as in Section 2.3, the field $\boldsymbol{w}^\ell$ that is refined within this minimization process is defined according to the minimizer at previous level $\ell+1$: $\boldsymbol{w}^\ell = \boldsymbol{w}^{\ell+1} + \Phi^{\ell+1}(\widehat{\mathsf{d}\boldsymbol{w}}^{\ell+1})$.[7] Similarly, the final segmentation $\hat{\mathcal{R}}^{\ell+1}$ at previous level $\ell+1$ is naturally used to initialize the segmentation process at level ℓ through its "projection" $[\Psi^\ell]^{-1} \circ \Psi^{\ell+1}(\hat{\mathcal{R}}^{\ell+1})$ which is well defined since the nestedness of the grids $\mathcal{B}^\ell$ implies the nestedness of the segmentation subsets Υ^ℓ ($\Upsilon^{\ell+1} = \mathrm{Im}\Psi^{\ell+1} \subset \Upsilon^\ell = \mathrm{Im}\Psi^\ell$).

As with motion estimation alone, the new multigrid function $\mathcal{H}^\ell$ turns out to be composed of four terms similar to those of $\mathcal{H}$: $\mathcal{H}^\ell = \mathcal{H}_1^\ell + \mathcal{H}_2^\ell + \mathcal{H}_3^\ell + \mathcal{H}_4^\ell$. The detailed expression of these different terms is provided in Appendix B.

3.3. Alternate Minimization

The joint minimization defined in Eq. (27) is conducted in turn w.r.t. the dense motion estimation variables (the increment $\mathsf{d}\boldsymbol{w}$ and the auxiliary weight sets β and δ) and the segmentation variables (the partition $\mathcal{R}$, the region-based motion parameters $\boldsymbol{\varphi}$, and the auxiliary variable set η). We now review the different minimizations to be alternated.

3.3.1. Dense Motion Field Update (Free Variables: $\mathsf{d}\boldsymbol{w}^\ell, \beta, \delta$; Fixed Variables: $\mathcal{R}^\ell, \boldsymbol{\varphi}, \eta$). The reduced segmentation $\mathcal{R}^\ell$ and the parametric likeness weights η being fixed, one has to solve:

$$(\widehat{\mathsf{d}\boldsymbol{w}}^\ell, \hat{\beta}, \hat{\delta}) = \arg\min_{(\mathsf{d}\boldsymbol{w}^\ell, \beta, \delta)} \left[\mathcal{H}_1^\ell + \mathcal{H}_2^\ell + \mathcal{H}_3^\ell\right]. \quad (28)$$

Apart from the interaction term $\mathcal{H}_3^\ell$, this is the same problem as in Section 2.3, and one can again resort to iteratively reweighted least squares. The different steps of this alternate optimization are as follows.

- The increment field $\mathsf{d}\boldsymbol{w}^\ell$ being fixed, let $\langle s, r\rangle \in \mathcal{C}$ be a pair of neighboring pixels and denote by m and n the block numbers (possibly identical) such that $s \in \mathcal{B}_m^\ell$ and $r \in \mathcal{B}_n^\ell$. From the β_{sr}'s point of view, the only change with respect to motion estimation case (17–18) occurs if $\langle s, r\rangle$ straddles two neighboring regions of $\Psi^\ell(\mathcal{R}^\ell)$. The optimal value of the discontinuity weights is given by:

$$\begin{aligned}&\forall i, \forall \langle s, r\rangle \subset \mathcal{R}_i, \beta_{sr}\\ &\quad = \frac{1}{\tau_2}\phi_2'\left[\left\|\left(\boldsymbol{w}_s^\ell + \mathsf{d}\boldsymbol{w}_n^\ell\right) - \left(\boldsymbol{w}_r^\ell + \mathsf{d}\boldsymbol{w}_m^\ell\right)\right\|^2\right], \end{aligned} \quad (29)$$

$$\begin{aligned}&\forall \langle i, j\rangle, \forall \langle s, r\rangle \in \partial\mathcal{R}_{ij}, \beta_{sr}\\ &\quad = \frac{1}{\tau_2}\phi_2'\left[\left\|\left(\boldsymbol{w}_s^\ell + \mathsf{d}\boldsymbol{w}_n^\ell\right) - \left(\boldsymbol{w}_r^\ell + \mathsf{d}\boldsymbol{w}_m^\ell\right)\right\|^2 \right.\\ &\qquad \left. + \frac{\mu_1}{\tau_2|\partial\mathcal{R}_{ij}|}\right]. \end{aligned} \quad (30)$$

For a pixel pair in between two neighboring regions, (30) implies that the optimal value is decreased as compared to the segmentation-free case (17–18) due to the shift by $\frac{\mu_1}{\tau_2|\partial\mathcal{R}_{ij}|}$ in the argument of the decreasing function ϕ_2'. The compound energy thus favors low discontinuity weights along the border of current segmentation.

- The data weights δ being only involved in $\mathcal{H}_1^\ell$, the update rule directly stems from (16) with piece-wise constant parameterization:

$$\begin{aligned}&\forall n \in S^\ell, \forall s \in \mathcal{B}_n^\ell, \delta_s\\ &\quad = \frac{1}{\tau_1}\phi_1'\left[\left(\boldsymbol{\nabla} f\left(s + \boldsymbol{w}_s^\ell, t+1\right)^T \mathsf{d}\boldsymbol{w}_n^\ell + f_t\left(s, \boldsymbol{w}_s^\ell\right)\right)^2\right]. \end{aligned} \quad (31)$$

- When the weights β and δ are frozen, the energy function $\mathcal{H}^\ell$ is quadratic with respect to $\mathsf{d}\boldsymbol{w}^\ell$. Its minimization is equivalent to the resolution of a linear system which is very similar to the one obtained with the hierarchical estimation of motion alone (with constant model). The only change comes from the influence of the segmentation-based parametric field to which the dense increment field is related within $\mathcal{H}_3^\ell$. See Appendix B.

3.3.2. Segmentation Update (Free Variables: $\mathcal{R}, \boldsymbol{\varphi}, \eta$; Fixed Variables: $\mathsf{d}\boldsymbol{w}, \beta, \delta$). The minimization of $\mathcal{H}^\ell$ w.r.t. the unknown segmentation $\mathcal{R}^\ell$ and associated parameters $\boldsymbol{\varphi}$, and w.r.t. the parametric likeness weights η, is conducted in the same alternate minimization spirit.

- First the segmentation is fixed, and the weights η_s's and the motion parameters $\boldsymbol{\varphi}_i$'s are estimated using iterated reweighted least squares. For a given region $\mathcal{R}_i^\ell \in \mathcal{R}^\ell$, the update of the motion parameter vector results from least squares regression

$$\begin{aligned}\boldsymbol{\varphi}_i = &\left[\sum_{n\in\mathcal{R}_i^\ell}\sum_{s\in\mathcal{B}_n^\ell} \eta_s P(s)^T P(s)\right]^{-1}\\ &\times \sum_{n\in\mathcal{R}_i^\ell}\sum_{s\in\mathcal{B}_n^\ell} \eta_s P(s)^T\left(\boldsymbol{w}_s^\ell + \mathsf{d}\boldsymbol{w}_n^\ell\right), \end{aligned} \quad (32)$$

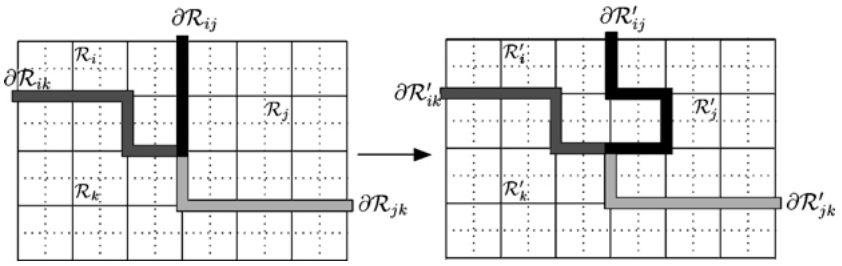


Figure 3. Local deformation of the segmentation: Example of local update changing segmentation $\mathcal{R} = \Psi^\ell(\mathcal{R}^\ell)$ into $\mathcal{R}' = \Psi^\ell(\mathcal{R}'^\ell)$ (with $\ell = 2$), by passing one boundary block of region $\mathcal{R}_i$ in adjacent region $\mathcal{R}_j$.

while the parametric likeness weights are updated according to:

$$\forall n \in \mathcal{R}_i^\ell, \forall s \in \mathcal{B}_n^\ell, \eta_s = \frac{1}{\tau_3}\phi_3'\left(\left\|\boldsymbol{w}_s^\ell + \mathsf{d}\boldsymbol{w}_n^\ell - P(s)\boldsymbol{\varphi}_i\right\|^2\right). \quad (33)$$

- Afterwards, the segmentation $\mathcal{R}^\ell$ is updated using a two-step process. First global changes of the segmentation (merging regions and creating new ones) are conducted and then local changes of the segmentation frontiers are considered.

The local updates consist in moving each point of the border $\partial\mathcal{R}^\ell$ within a small neighborhood such as to make the energy decrease. Let $\mathcal{R}'^\ell$ be the candidate modified segmentation (Fig. 3). Assuming that the adjacency graph of the segmentation remains the same in this local deformation, the associated energy variation is:

$$\begin{aligned}\mathcal{H}^\ell(\mathcal{R}'^\ell) - \mathcal{H}^\ell(\mathcal{R}^\ell) &= \lambda 2^\ell(|\partial\mathcal{R}'^\ell| - |\partial\mathcal{R}^\ell|) \\ &+ \sum_{\langle i,j\rangle}\left[\frac{\mu_1}{\left|\partial\mathcal{R}_{ij}^\ell\right|}\sum_{\langle s,r\rangle\in\partial\mathcal{R}_{ij}^\ell}\beta_{sr} - \frac{\mu_1}{\left|\partial\mathcal{R}'^\ell_{ij}\right|}\sum_{\langle s,r\rangle\in\partial R'^\ell_{ij}}\beta_{sr}\right] \\ &+ \sum_i\left[\sum_{n\in\mathcal{R}'_i\setminus\mathcal{R}_i}\sum_{s\in\mathcal{B}_n^\ell}\phi_3\left(\left\|\boldsymbol{w}_s^\ell + \mathsf{d}\boldsymbol{w}_n^\ell - P(s)\boldsymbol{\varphi}_i\right\|^2\right)\right. \\ &\left. - \sum_{n\in\mathcal{R}_i\setminus\mathcal{R}'_i}\sum_{s\in\mathcal{B}_n^\ell}\phi_3\left(\left\|\boldsymbol{w}_s^\ell + \mathsf{d}\boldsymbol{w}_n^\ell - P(s)\boldsymbol{\varphi}_i\right\|^2\right)\right]. \quad (34)\end{aligned}$$

This local energy variation is easily computed. In practice, a new position is considered for each border element of the current segmentation $\mathcal{R}^\ell$. If this position corresponds to an energy decrease it is accepted and the map is updated. In our experiments, a border element is allowed to move one site forward or backward in the direction perpendicular to the border. Let us note that these displacements may be quite large since they actually correspond to 2^ℓ pixels. Therefore, the optimal motion parameters associated with the new region $\mathcal{R}_i$ may substantially changed and should be ideally re-estimated. Instead of such a joint update of region geometry and parameterization, we chose a cheaper alternate minimization: we keep all the motion parameters fixed during a complete visit (one iteration) of all the boundary elements; then the different motion parameters are re-estimated at the same time for the whole segmentation. Each step makes the global energy decrease, which guarantees the convergence to a local minima.

Beside local deformations, global updates allow to change at once a whole region as well as the topology of the segmentation (number and connectivity of regions). In this work we only consider global transformations based on the creation of new regions and the merging of adjacent regions.

The merging of two adjacent regions consists in removing their common boundary, when this yields a global energy decrease. $\mathcal{R}^\ell$ being the current segmentation, the energy variation associated with the merging of two regions i and j is derived in the same way as for the local deformation step, except that one has to compute the new parametric model associated with $\mathcal{R}_i^\ell \cup \mathcal{R}_j^\ell$ to determine the actual energy of the new segmentation. This is done for each pair of adjacent regions. The boundary leading to the greatest energy decrease is removed and the corresponding regions are merged. This process is repeated until a complete stability is reached.

The inclusion of a new region could be done at random. It is much more effective to devise a data-driven mechanism based on a simple assessment of the locations whose dynamic content is not well explained by the current segmentation. We propose to conduct this

search of good candidates for new regions by classifying the parametric likeness weights η_s's within two classes (outliers and inliers) according to both their individual values and those carried by their neighbors. A classic way of achieving such a contextual image classification is to minimize w.r.t. labeling $x^\ell = \{x_n^\ell \in \{0, 1\}, n \in \mathcal{B}^\ell\}$ the following cost function:

$$\sum_{n\in S^\ell}\sum_{s\in\mathcal{B}_n^\ell}\frac{1}{2\sigma^2\left(x_n^\ell\right)}\left[m\left(x_n^\ell\right)-\eta_s\right]^2 + \log\sigma\left(x_n^\ell\right)+\sum_{\langle n,m\rangle}\delta\left(x_n^\ell, x_m^\ell\right), \tag{35}$$

where $\delta(\cdot)$ is the delta Kronecker function. The parameters $\sigma(0)$ and $m(0)$ (respectively $\sigma(1)$ and $m(1)$) stand for the standard deviation and the mean of the outlier class (respectively the inlier class). The values of these parameters have been learned on typical examples. Based on this learning, they were fixed to $m(0) = 0.05$, $m(1) = 0.98$, and $\sigma(0) = 0.5$, $\sigma(1) = 0.05$ in all the experiments. The minimizer is searched with the deterministic ICM algorithm (Besag, 1986). If a sufficiently large region of connected outliers is recovered (at least 64 pixels in our experiments) and if turning this set into a new region yields an energy decrease, the new region is effectively incorporated within the current segmentation.

Global deformations obviously involve far more computations than local deformations. In practice, we only use the global transformations at the beginning of each level ℓ. They provide quickly sensible segmentations which are then only refined through frontiers updates.

Concerning the global procedure, let us recall that it is multiresolution (based on a pyramid of images), and hierarchical within each resolution. As concerns the initialization at the coarser resolution, the initial motion field is set to zero and associated with a partition composed of a unique region (the whole image). Since we face a non-convex minimization problem, the results depend on the initialization. Nevertheless as will be shown in the experiments of Section 4, we did not observe a strong dependency to the initialization in practice, and the choice of the simple initialization above turned out to yield consistently satisfactory results. As for the global convergence of the method, the dense motion estimation process and the motion segmentation process that are alternated have to be discussed separately. Since the computation of the motion field is performed through the hierarchical Gauss-Newton minimization presented in Section 2.4, it exhibits the convergence properties and limitations already discussed. At fixed motion field, each of the elementary minimizations w.r.t. the different ingredients of the segmentation, leads to a decrease of the global energy. This guarantees that this part of the whole procedure converges toward a local minima. Again, although the complete method may fail to converge due to the Gauss-Newton nature of the dense motion computation, we never observed in practice divergent sequences of iterates, even within the intensive tests conducted on synthetic data for a wide range of parameter values.

3.4. *Continuous Point of View*

As opposed to the dense motion estimation alone, the joint estimation-segmentation is not easy to express within a continuous formalism. The reason is twofold.

First, it is a complicated issue in its own to simply specify and manipulate partitions of the continuous image plane. A general and rigorous definition implies the use of an unknown number of pieces of Jordan curves, connected at junction points (Mumford and Shah, 1989). Another way consists in using a known number of interacting level sets (Samson et al., 1999; Yezzi et al., 1999). Both approaches are complex and lack flexibility. Hence, it seems to be no continuous counterpart to labeling the discrete pixel grid into different regions.

We can nevertheless write down a continuous analog of the interaction term (23) as:

$$\mu_1\sum_{\langle i,j\rangle}\int_{\partial\mathcal{R}_{ij}}\beta(\boldsymbol{x}(s))\,ds+\mu_2\sum_i\iint_{\mathcal{R}_i} \times\underbrace{\left[\tau_3\eta(\boldsymbol{x})\|\boldsymbol{w}(\boldsymbol{x})+\mathsf{d}\boldsymbol{w}(\boldsymbol{x})-P(\boldsymbol{x})\boldsymbol{\varphi}_i\|^2+\psi_3(\eta(\boldsymbol{x}))\right]}_{\triangleq F_i(\boldsymbol{x})}d\boldsymbol{x}. \tag{36}$$

The Euler-Lagrange equation for the compound functional provides conditions on functions β and η which are the continuous counterpart of the discrete update rules (29–30) and (33). Besides, changing the second surface integral in (36) into a curvilinear one with Green theorem (as in Zhu and Yuille (1996), shows that each boundary $\partial\mathcal{R}_{ij}$ evolves as a snake driven by an internal force which depends on the chosen prior, and an external force:

$$-\boldsymbol{\nabla}\beta(\boldsymbol{x})+(F_j(\boldsymbol{x})-F_i(\boldsymbol{x}))\boldsymbol{n}_{ij}(\boldsymbol{x}), \tag{37}$$

where $\boldsymbol{n}_{ij}$ is the normal to the frontier $\partial\mathcal{R}_{ij}$ directed toward the interior of $\mathcal{R}_i$.

Our interaction mechanism based on the discontinuity auxiliary variables could as well be considered in case region borders are implicitly defined in terms of level sets. The mechanism we introduce thus offers a new way to deal with the generic issue of joint anisotropic diffusion and segmentation, in both discrete and continuous frameworks.

The second reason for which it seemed more natural to let our estimation-segmentation approach within the discrete realm is related to the choice of the minimization technique. Using segmentations of the discrete pixel grid enables us to extend in a very simple way the efficient hierarchical minimization scheme previously introduced for the estimation alone.

4. Experimental Results

In this section we present results of dense motion estimation with and without joint motion-based segmentation. In both cases, quantitative comparisons are presented on a synthetic sequence. As for experiments on real data, the segmentation-free motion estimation is demonstrated on sequences that seem not to admit any simple region-wise parametric description. The adaptive compromise between local dense representation and more global parametric one which is offered by the hierarchical minimization allows us to get interesting results on fluid sequences for instance. At the opposite, the parametric description used in the joint estimation-segmentation approach is more suited to scenes that involve mostly rigid objects.

Note that in both approaches we selected Leclerc estimator, $\phi(x^2) \triangleq 1 - \exp(-\tau^{-1}x^2)$, for the different robust functions (Leclerc, 1989). Also, in both cases we chose the following stopping criterion for the iterative estimation scheme at a given resolution/grid level:

$$\#\left\{s \in S : \frac{\|\mathsf{d}\boldsymbol{w}_s(n+1) - \mathsf{d}\boldsymbol{w}_s(n)\|}{\|\mathsf{d}\boldsymbol{w}_s(n+1)\|} < 0.01\right\} < \varepsilon, \tag{38}$$

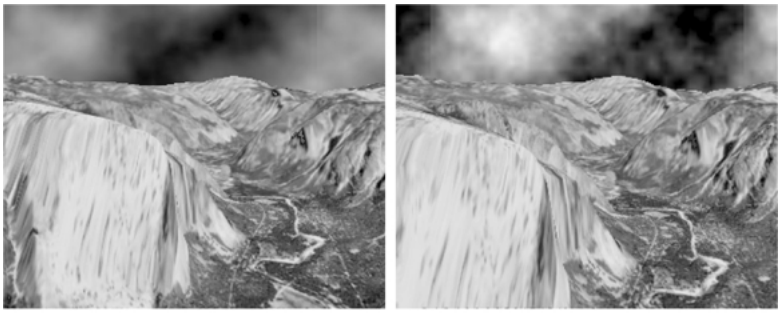
where $\mathsf{d}\boldsymbol{w}_s(n)$ denotes the iterate at step n and point s. In the motion estimation experiments ε has been fixed to 1% of the number of points, whereas for the joint segmentation-estimation method, a value of 5% was sufficient to improve the accuracy of the motion field on a synthetic benchmark.

4.1. Results of Dense Motion Estimation

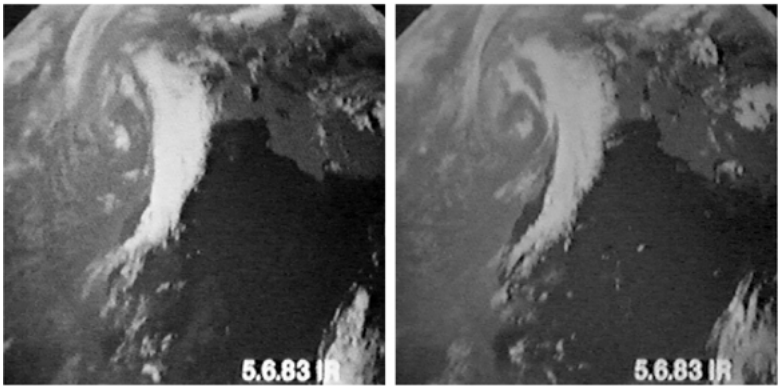
The experiments have been carried out both on a synthetic sequence—for which a ground truth is known and comparative quantitative comparisons can be reported (Barron et al., 1994)—and on real world sequences (Fig. 4). The synthetic sequence is the well known *Yosemite* sequence used in the comparative benchmark of Barron et al. (1994). The real data are composed of two particularly challenging sequences which involve the highly deformable motion of fluid media. The first one, named *Depression*, is a meteorological video sequence involving large displacements. It includes a through of low pressure and different moving clouds. The second one, named *Smoke*, comes from fluid mechanics experiments with smoke undergoing complex motion under poor lighting conditions.

The values of the parameters that define the hierarchical structure in these experiments were set as follows: the number of resolution levels was respectively 2 for *Yosemite* and *Smoke*, and 3 for *Depression*. The number of grid levels was fixed to 6 for *Smoke* and to 5 for the two others.

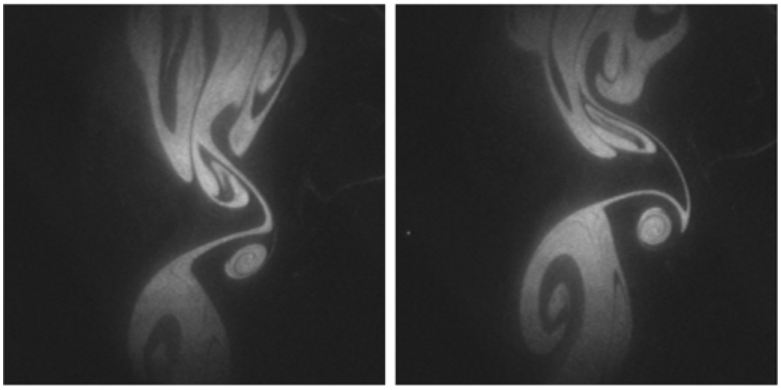
As for the hierarchical constrained minimization, both *regular* and *adaptive* partitions into square patches have been considered. In the regular case, the partition $\mathcal{B}^\ell$ is composed of $2^\ell \times 2^\ell$ square blocks and the associated adjacency graph is a regular lattice with the same neighborhood system as the original grid. In the adaptive case, the partition $\mathcal{B}^{\ell-1}$ is determined online, based on the previous partition $\mathcal{B}^\ell$ and on the associated final estimate $(\hat{\boldsymbol{\theta}}^{\,\ell}, \hat{\beta}^{\,\ell}, \hat{\delta}^{\,\ell})$. The new partition is obtained by dividing some of the elements of $\mathcal{B}^\ell$ according to a splitting criterion to be defined. It seems natural to base this criterion upon the agreement of the current motion estimate with the luminance conservation assumption, measured on block $\mathcal{B}_n^\ell$ by $\sum_{s\in\mathcal{B}_n^\ell}[f(s + \boldsymbol{w}_s^\ell + P_n(s)\hat{\boldsymbol{\theta}}_n^{\,\ell}, t+1) - f(s,t)]^2$ (sum of squared registration errors), or, in linearized form: $\sum_{s\in\mathcal{B}_n^\ell}[\nabla f(s+\boldsymbol{w}_s^\ell, t+1)^T P_n(s)\hat{\boldsymbol{\theta}}_n^\ell + f_t(s, \boldsymbol{w}_s^\ell)]^2$. Instead of using this quantity which has to be computed, we use the final data weights $\hat{\delta}_s$'s, which are function of the squares in the sum above according to (16). Experimental evidence indicated that it is more appropriate to consider how uniform (instead of how good) is the quality of the agreement within considered patch. A block is thus divided into four sub-blocks if the standard deviation of $\{\hat{\delta}_s^{\,\ell}, s \in \mathcal{B}_n^\ell\}$ exceeds a given threshold. In all the experiments this threshold is set to 0.05.



(a) *Yosemite*



(b) *Depression*



(c) *Smoke*

Figure 4. Sequences used in dense motion estimation experiments: (a) frames 0 and 14 of synthetic sequence *Yosemite*, size 288×224; (b) frames 1 and 9 of satellite sequence *Depression*, size 256×256; (c) frames 10 and 14 of experimental fluid mechanics sequence *Smoke*, size 512×512.

As mentioned in Section 2.4, two different parameterizations corresponding respectively to 2 and 6 parameters are considered. We use them within three different combinations denoted M_6, M_2, and M_{62} where subscripts indicate allowed parameterizations. Models M_6 and M_2 deal with a single type of parameterization. In these two cases the hierarchical minimization is stopped when a certain minimal size (8×8, and 1×1 resp.) is reached by the smallest patches of the current partition. In contrast, model M_{62} mixes different parameterizations in the following way: the affine model is used for all blocks at least as large as 8×8, and the

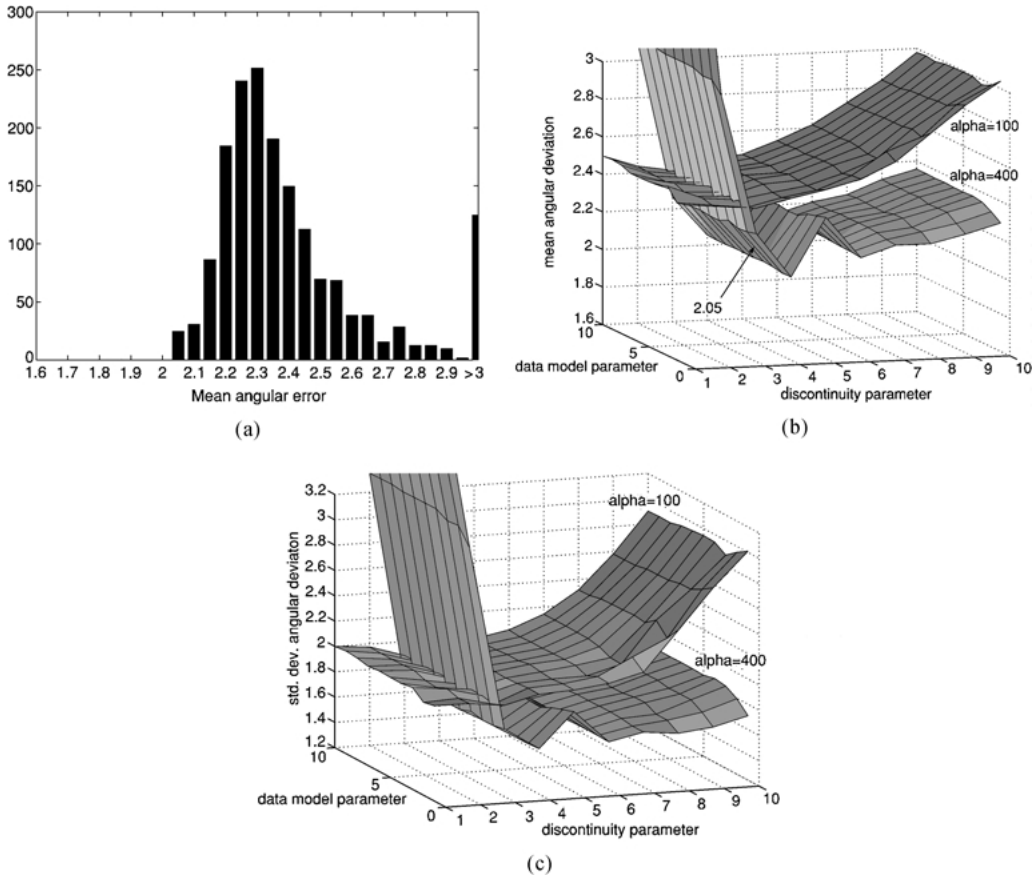


Figure 5. Dense motion estimation on *Yosemite* with M_6 (affine constraint in the hierarchical minimization) and regular partitioning: (a) histogram of the mean angular discrepancy for 1700 parameter triples; (b) surfaces of the mean angular error for $\alpha = 100$ and $\alpha = 400$; (c) surfaces of the std. dev. of the angular error for $\alpha = 100$ and $\alpha = 400$.

constant model is used for blocks of size 4×4 and less.

Following Barron et al. (1994), we provide quantitative comparative results on *Yosemite*. For each estimate, the angular deviation with respect to the real flow is computed at "reliable" locations (the percentage of such locations is the "density" of the estimate; it is 100% in our case). Let us note that in order to stick to the experimental conditions reported in the best recent studies on that sequence we consider only a cropped sequence where the sky was removed. To assess the performance of our motion estimation technique on this sequence, we have compared its three versions M_6, M_2, and M_{62}, with regular partitioning, and for 1700 different parameter triples $(\alpha, \tau_1^{-1}, \tau_2^{-1}) \in [100, 420] \times [2, 36] \times [0.2, 1.1]$ with sampling steps 20, 4, and 0.1, respectively in each direction.

Figures 5–7(a) show the histograms of the mean angular discrepancies obtained. Beside, we show also for two extreme values of the smoothness parameter the surfaces corresponding respectively to the mean and the standard deviation of the angular discrepancy, the two other parameters varying according to ranges previously indicated. The best parameter combinations are then compared with the adaptive partitioning strategy.

In the case of regular grid partitioning, the M_6 model performs less well than the two other versions. Although slightly less stable the constant model within M_2 yields very good results. For a slightly increased cpu time, the best results are provided by the mixed model embedded in M_{62}, as can be seen by comparing the three histograms in Figs. 5–7(a) for values exceeding 3° (i.e., last bar). Histogram 7(a) indicates that with M_{62} almost 60% of the trials fall

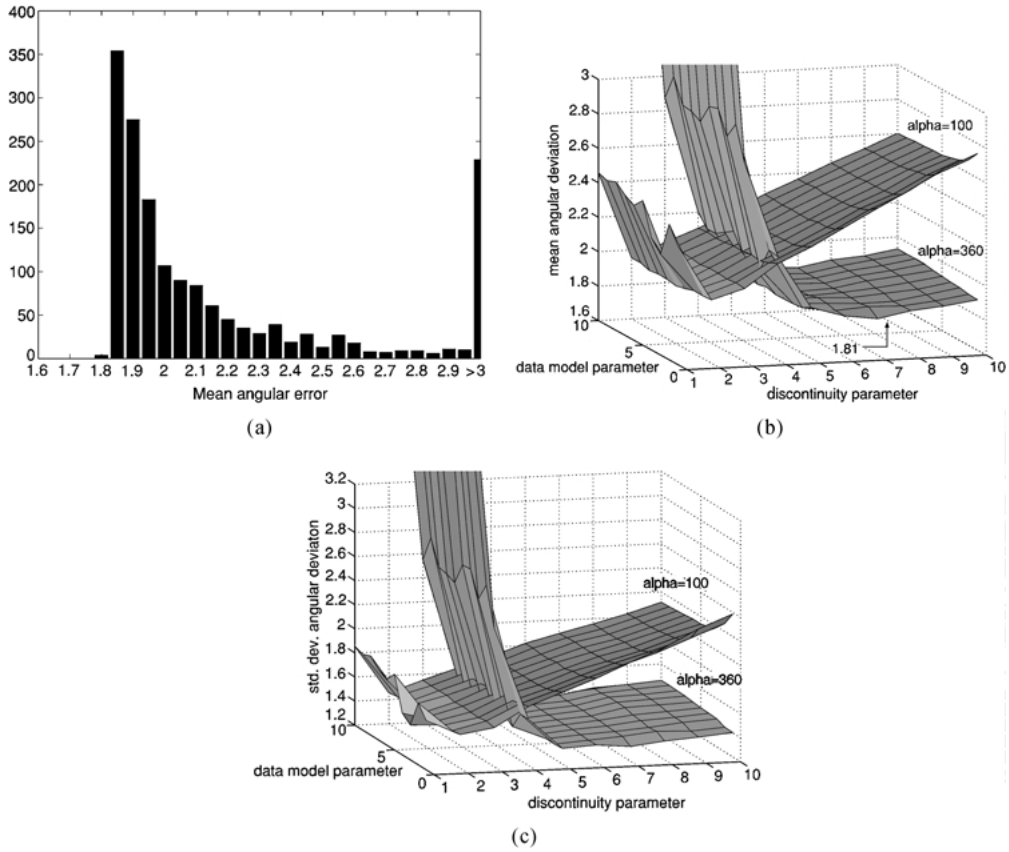


Figure 6. Dense motion estimation on *Yosemite* with M_2 (constant constraint in the hierarchical minimization) and regular partitioning: (a) histogram of the mean angular discrepancy for 1700 parameter triples; (b) surfaces of the mean angular error for $\alpha = 100$ and $\alpha = 360$; (c) surfaces of the std. dev. of the angular error for $\alpha = 100$ and $\alpha = 360$.

below an average angular error of 2 degrees. To outline the quality of these results, we superimposed on the histogram in Fig. 7(a) the best results reported in the state-of-the-art literature. The proposed method is able to outperform these other techniques for a wide range of variations of its three parameters. It is also interesting to notice in the surface plots in Figs. 5–7(b) and Figs. 5–7(c) the low sensitivity of the method with respect to the parameter τ_1 of the data robust penalty.

For the three different versions of our model, as well as for the two types of partitioning strategies, the best mean angular error (μ), the associated standard deviation (σ), and the set of parameters for which they have been obtained are gathered in Table 1. This table also lists the corresponding

Table 1. Best results on *Yosemite* for the three versions of our approach, along with associated parameters and cpu times.

		Regular partition			Adaptive partition		
Model	Parameters	μ	σ	cpu	μ	σ	cpu
M_6	$\alpha = 400,\ \tau_1 = 3.0,\ \tau_2 = 0.3$	2.04°	1.57°	63 s	1.81°	1.34°	46 s
M_2	$\alpha = 360,\ \tau_1 = 4.5,\ \tau_2 = 0.2$	1.81°	1.33°	145 s	2.46°	1.74°	41 s
M_{62}	$\alpha = 360,\ \tau_1 = 3.0,\ \tau_2 = 0.8$	1.73°	1.33°	199 s	1.93°	1.33°	65 s

Table 2. Comparative results on *Yosemite*: the first part of the table corresponds to results reported by Barron et al. (1994), where references to the compared methods are to be found; the second part gathers results reported more recently in the referenced literature, on a cropped sequence without sky.

Technique	μ	σ	density
Horn and Schunck (original)	31.69°	31.18°	100%
Horn and Schunck (modified)	9.78°	16.19°	100%
Uras et al.	8.94°	15.61°	100%
Lucas and Kanade	4.28°	11.41°	35.1%
Fleet and Jepson	4.63°	13.42°	34.1%
Without sky			
Bab-Hadiashar and Suter (1998)	1.97°	1.96°	100%
Lai and Vemuri (1998)	1.99°	1.41°	100%
Ju et al. (1996)	2.16°	2.0°	100%
Black and Jepson (1996)	2.29°	2.25°	100%
Mémin and Pérez (1998a)	2.34°	1.45°	100%
Szeliski and Coughlan (1994)	2.45°	3.05°	100%
Black (1994)	3.52°	3.25°	100%

cpu times measured on a Sun Ultra Sparc (200 Mhz).

In order to give elements of comparison, Table 2 recalls some of the results presented by Barron et al. (see corresponding references therein). They concern an adaptation of Horn and Schunck's algorithm, the best full-density algorithm (Uras et al.) and the two algorithms yielding the best results, but with reduced densities (Lucas and Kanade, Fleet and Jepson). In the lower part of the table, we include the results obtained by other authors on the same sequence with sky removed.

It appears in Table 1 that the adaptive partitioning provides a noticeable speed up as compared to the regular partitioning. Except for M_6, this acceleration is obtained at the cost of a slight loss of quality, at least on that synthetic sequence and in view of the angular discrepancy criterion. At that point, we would like to stress out that this discrepancy measure used to assess the quality of motion fields should be interpreted with caution due to its global nature (it is just a mean). In particular this criterion does not

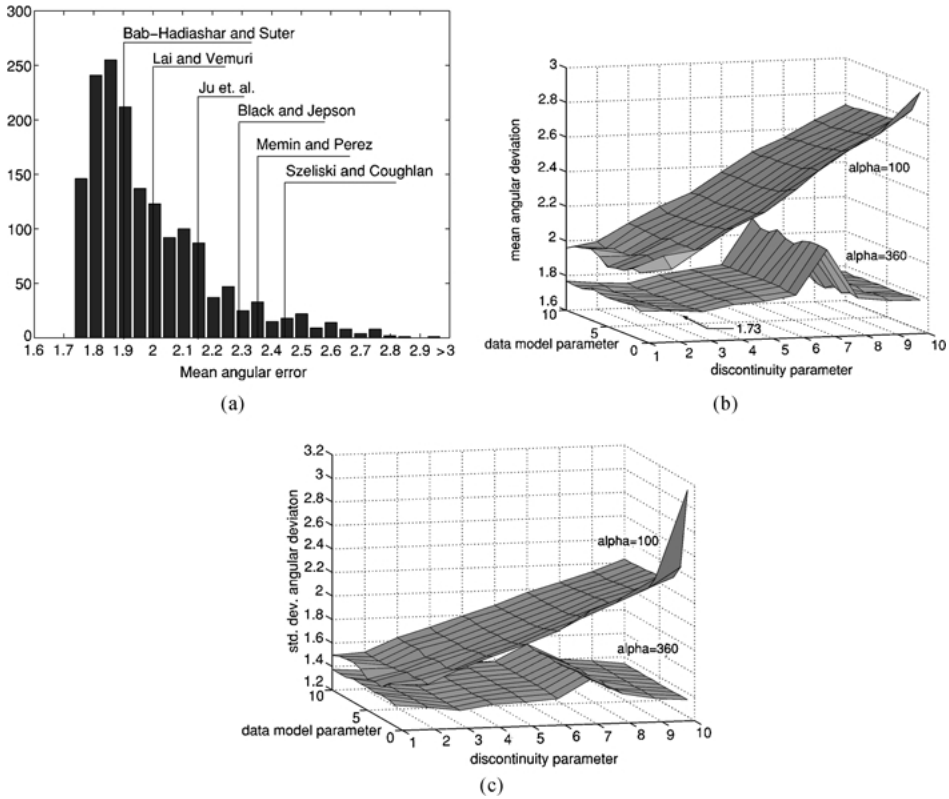


Figure 7. Dense motion estimation on *Yosemite* with M_{62} (mix of affine and constant constraints in the hierarchical minimization) and regular partitioning: (a) histogram of the mean angular discrepancy for 1700 parameter triples; (b) surfaces of the mean angular error for $\alpha = 100$ and $\alpha = 360$; (c) surfaces of the std. dev. of the angular error for $\alpha = 100$ and $\alpha = 360$.

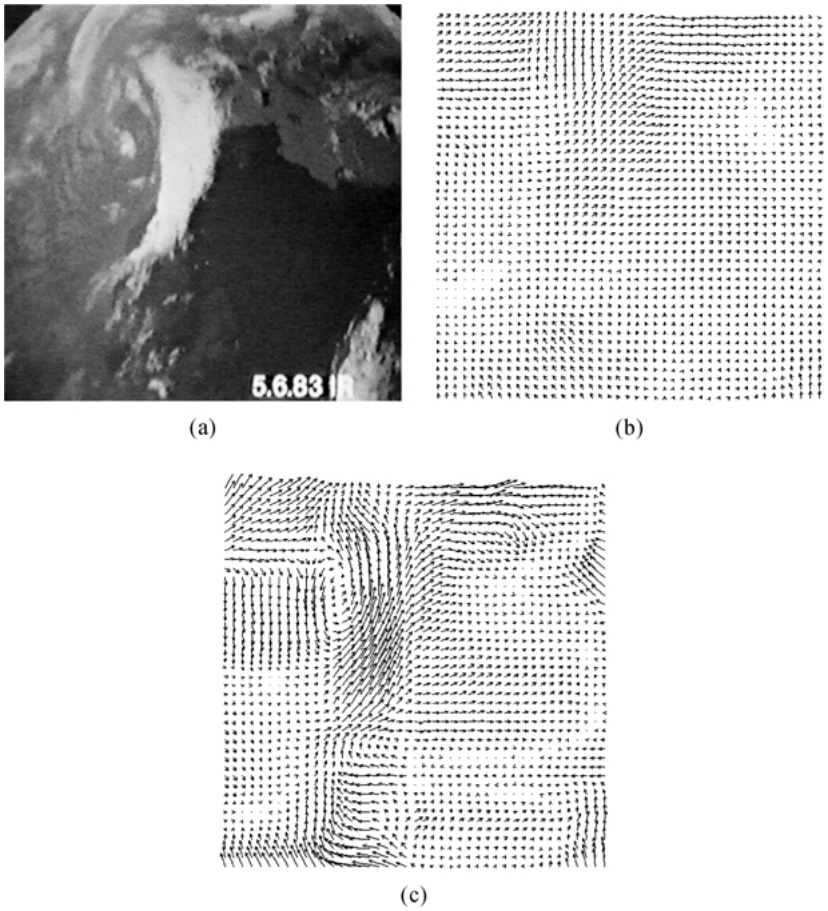


Figure 8. Results on satellite sequence *Depression*: (a) one frame; (b) motion field estimated with M_2 and regular partitioning (cpu time: 105 s); (c) motion field estimated with M_6 and adaptive partitioning (cpu time: 24 s).

allow to evaluate the critical ability of a particular method to locate and preserve discontinuities. For instance, the adaptive M_6 model, which gives better results "in average", is not able to estimate accurately the motion around spatial discontinuities-due to the crude partition on which it lies. The mixed model M_{62} with adaptive grids performs better from that point of view, which will prove useful in real world cases. Nevertheless one must keep in mind that the use of adaptive grids requires the tuning of a supplementary parameter.

Note that in the case of M_2 associated with a regular subdivision, the results are improved as compared to the pure top-down multigrid method we had introduced in Mémin and Pérez (1998a). Hierarchical Gauss-Newton with its successive inter-level warping performs better on this particular example.

We now turn to the real world sequences described at the beginning of the section. Figure 8 presents for *Depression* the final motion fields respectively estimated by M_2 with the regular division and M_6 with the adaptive division. The two vector fields are displayed the same way, namely sub-sampled by 6 and magnified by 4. We can notice that with the regular piecewise constant constraint on increments, the flow is drastically under-estimated and over-smoothed as compared to the one produced with the adaptive affine constraint. As a consequence, local features of interest such as the depression center in the left upper corner of the image are concealed with M_2 whereas, with M_6, this depression center is clearly visible and may be easily identified in an automatic way. This real example demonstrates that the use of the model M_6, as well as the use of the adaptive partitioning not

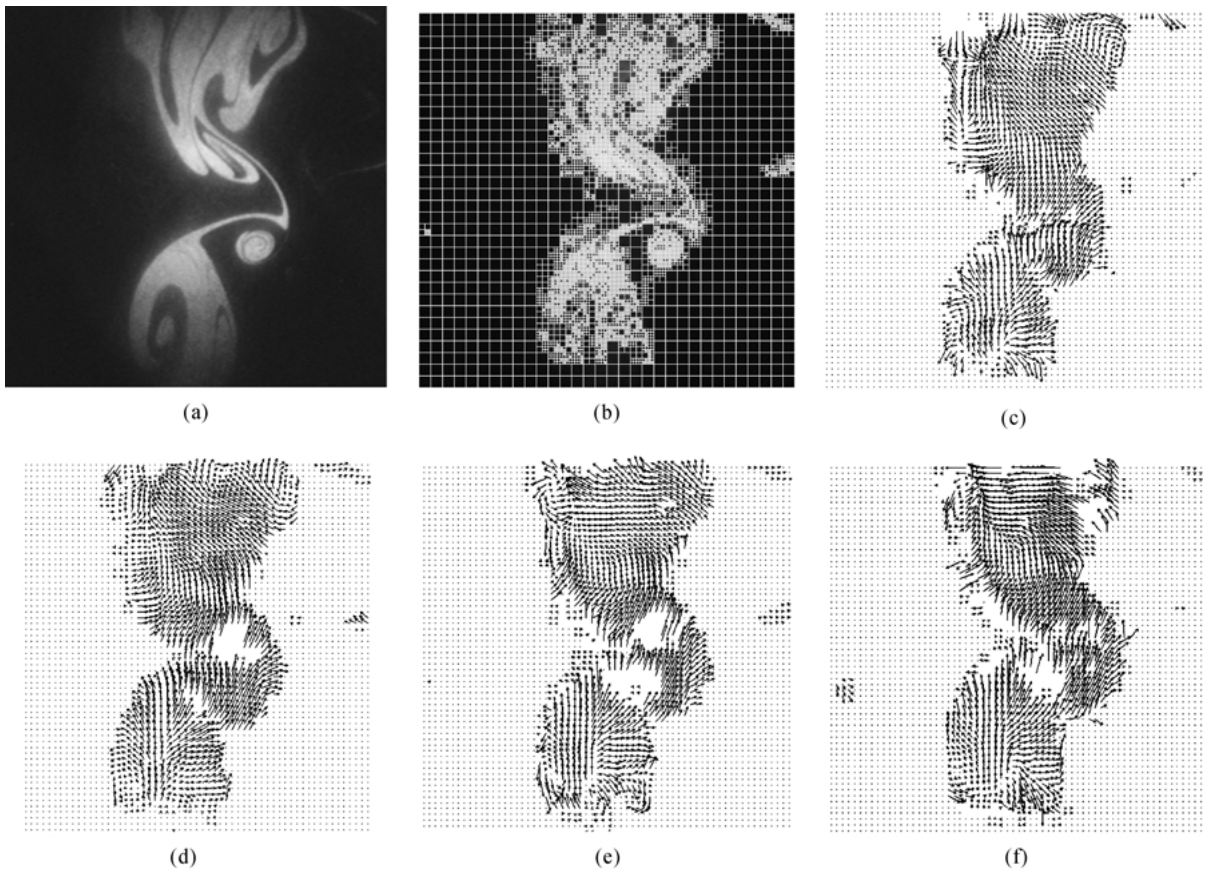


Figure 9. Results on experimental fluid mechanics video *Smoke*: (a) one frame; (b) final partition; (c–f) four consecutive motion fields estimated with M_{62} and adaptive partitioning.

only reduces the computational load but can also result in improved results, contrary to what we observed on *Yosemite*.

The sequence *Smoke* is all the more difficult an example since it implies large displacements (up to 20 pixels) and low photometric gradients. The final estimation partition obtained with M_{62} for one image of the sequence is given in Fig. 9(b) along with four consecutive motion fields estimated from the sequence. Visually, the estimates obtained seem compliant with the apparent dynamics of the fluid flow. In addition, the estimation turns out to be quite stable in time despite the absence of any temporal link.[8]

The whole multiresolution/hierarchical algorithm converges quickly, with only ten or so low cost iterations at each level. Also, as thoroughly assessed with the synthetic data, the technique exhibits a low sensitiveness to parameter values within large ranges of variation. This confirms the observations that have been made in Mémin and Pérez (1998a) for the constant model associated with the pure top-down multigrid method and in Hellier et al. (2001) for an extension of the adaptive M_6 model to the registration of volumic MR brain images (with the adaptive partitioning being in addition driven by anatomical structures). In the latter study the dedicated technique proved in addition capable to deal with complex inter-patient topological changes.

4.2. *Results of Joint Estimation-Segmentation*

In the case of the joint estimation-segmentation approach, we also report comparative results on *Yosemite* for comparison purpose. We then consider

(a) *Parking lot*

(b) *Calendar*

Figure 10. Sequences used in motion estimation-segmentation experiments: (a) frames 1 and 4 of outdoor sequence *Parking lot*, size 224×224; (b) frames 1 and 2 of indoor sequence *Calendar*, size 256×256.

two real-world sequences (Fig. 10). The first sequence is a *Parking lot* sequence which involves two cars moving in the foreground while the camera pans the scene. The second one, named *Calendar*, includes several moving objects (a calendar moving vertically and a toy train pushing a ball) and a horizontal panning of the camera.

As for the parameter values, the number of resolution levels was respectively 2 for *Calendar* and *Yosemite* and 1 for *Parking lot*. The number of grid levels was fixed to 6 for *Calendar* and to 5 for the two others. Most of the energy parameters were kept the same for the three sequences: $\alpha = 100$, $\tau_1^{-1} = 5$, $\tau_2^{-1} = 0.3$, $\tau_3^{-1} = 0.3$, $\mu_1 = 30$. We only made sequence dependent the parametric likeness parameter μ_2 and the segmentation *a priori* parameter λ.

To assess the performances and the stability of the joint estimation-segmentation method we run it on a set of 150 values of this couple (μ_2, λ) within $[10,\ 100] \times [2,\ 70]$. The sampling step is 10 for μ_2 and for λ it is 2 between 2 and 20 and 10 beyond. Figure 11 shows the surface of the obtained mean angular discrepancy and the corresponding histogram.

Table 3. Dense motion segmentation vs. joint estimation-segmentation on *Yosemite* with $\lambda = 12$ and $\mu_2 = 70$.

Model	μ	σ
Estimation-segmentation		
Parametric estimate	1.58°	1.21°
Dense estimate	1.92°	1.59°
Dense estimation only		
M_2 regular	2.91°	3.17°

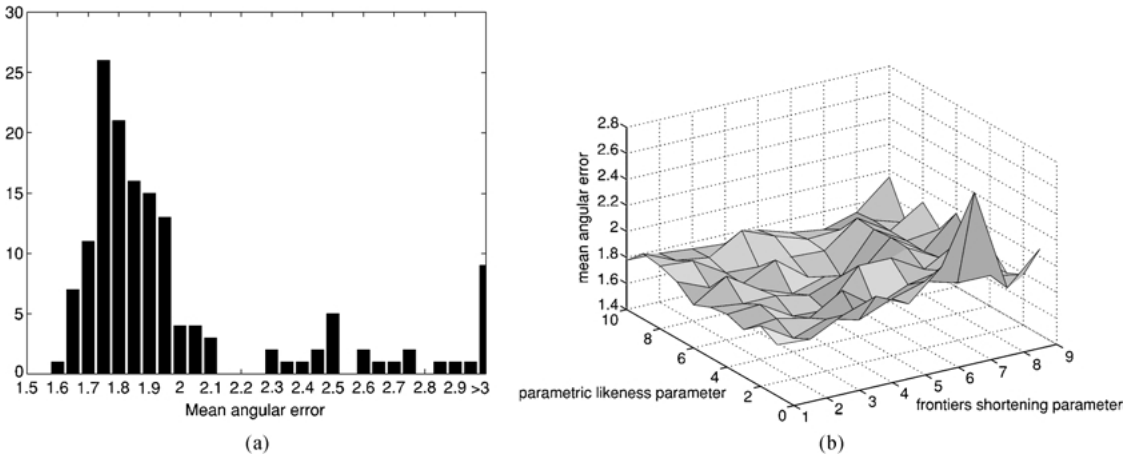


Figure 11. Joint segmentation-estimation on *Yosemite* for different parameter values: (a) histogram of the mean angular discrepancy for 150 parameter triples (μ_2, λ); (b) corresponding surface of the mean angular discrepancy.

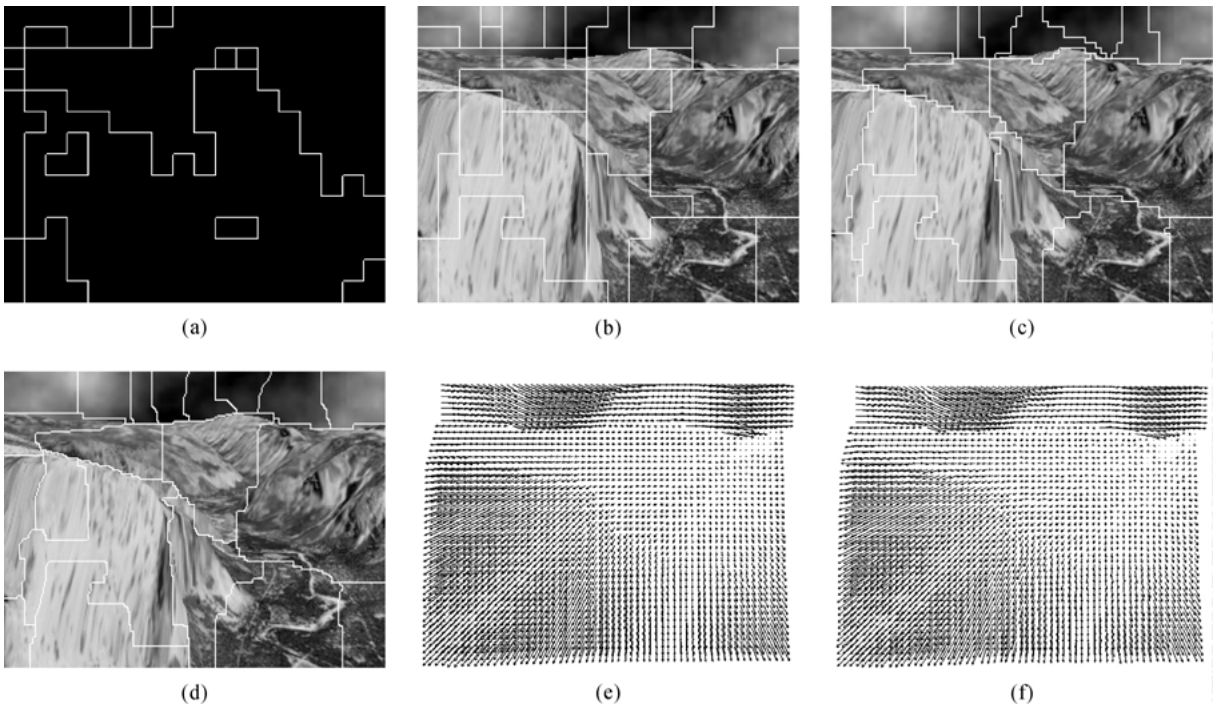


Figure 12. Joint motion estimation-segmentation on *Yosemite*: (a) segmentation initialization at the coarsest level $\ell = 4$; (b–d) final segmentations at grid levels $\ell = 4, 2, 0$; (e) final dense field estimate; (f) final parametric field estimate (cpu time $\sim$8 mn).

In Table 3, we compare the best results obtained on *Yosemite* by the joint estimation-segmentation method ($\lambda = 12$ and $\mu_2 = 70$) to those obtained with the dense estimation alone. The motion parameters and the convergence criteria (with $\varepsilon = 5\%$) were the same. The final fields obtained by the joint approach, as well as some successive segmentations are shown in Fig. 12. Both the parametric and dense motion fields jointly obtained significantly improve the mean angular error, as compared to the dense estimation alone. A dense field

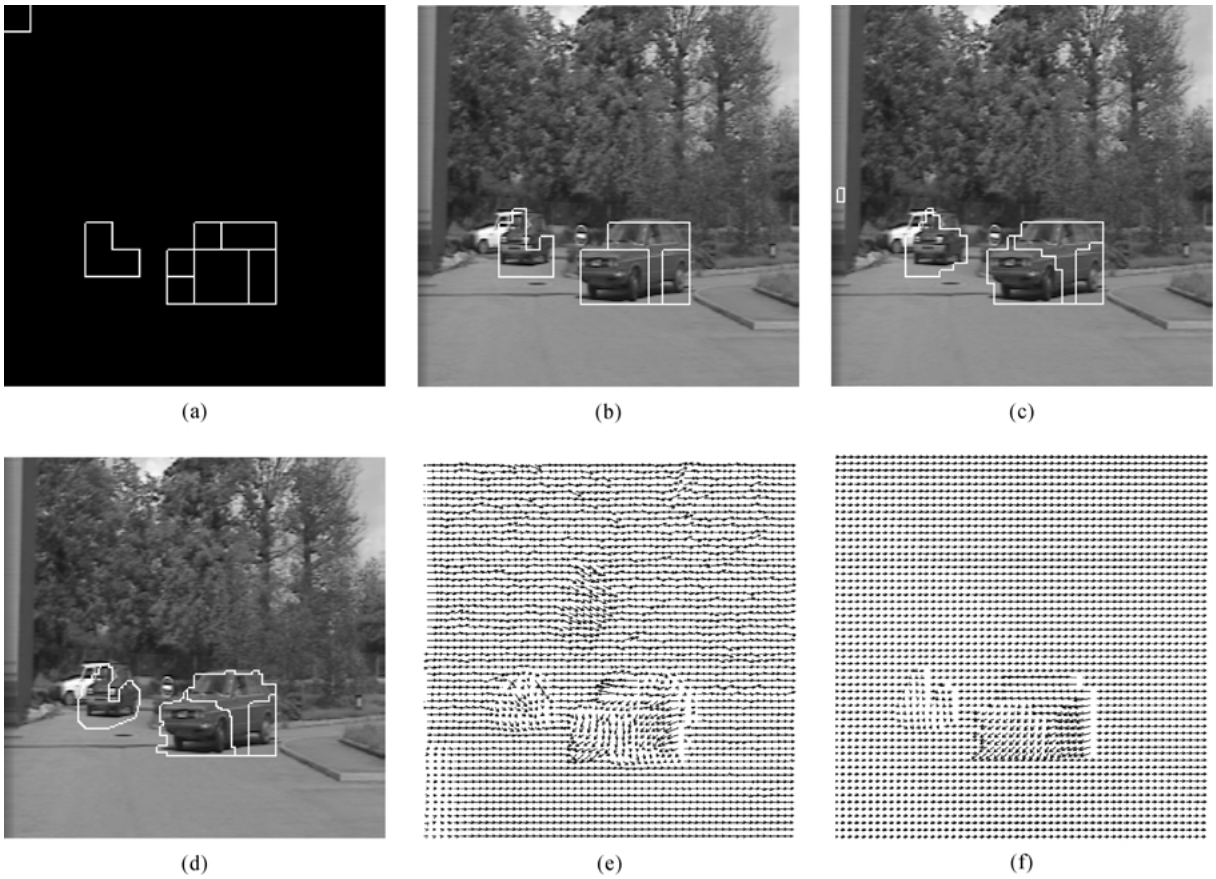


Figure 13. Joint motion estimation-segmentation on *Parking lot*: (a) segmentation initialization at level $\ell = 4$, (b–d) final segmentations at grid levels $\ell = 3, 2, 0$; (e) final dense field estimate; (f) final parametric field estimate (cpu time $\sim$4 mn).

of better quality is thus obtained along with a compact piecewise parametric field that approximates it very well. It is worth noting that this piecewise representation of the motion field further enhances the ability of the hierarchical dense motion estimator to extrapolate nicely over large occlusion areas. In *Yosemite*, the displacement of the region exiting on the left of the image plane due to the large divergent motion is particularly well recovered.

As for the segmentation itself, two things can be noticed. First, although the segmentation process only interacts indirectly with the data through the dense motion field under estimation, it is able to split the scene into pieces that make sense from the tri-dimensional point of view (i.e., the different image segments correspond to different motions, depths, or orientations). For instance the partitioning of the mountains in *Yosemite* and of the front car in *Parking lot* are consistent with the three-dimensional structures of these rigid objects. Secondly, the boundaries of the segmented regions fit correctly the discontinuities of the apparent motion. See for example the crest of both the foreground and the background mountains in *Yosemite*, the front car in *Parking lot*, and the train engine and the rolling ball in *Calendar*.

Finally, let us outline that the joint motion estimation-segmentation is not very sensitive to the initialization. As shown in Figs. 12 and 14, it is indeed able to recover meaningful partitions of the displacement field from rather poor initializations.

Rough estimates of the computation times (code not hand-optimized) obtained on a 200 Mhz Sun Ultra Sparc are also given in the captions of Figs. 12–14.

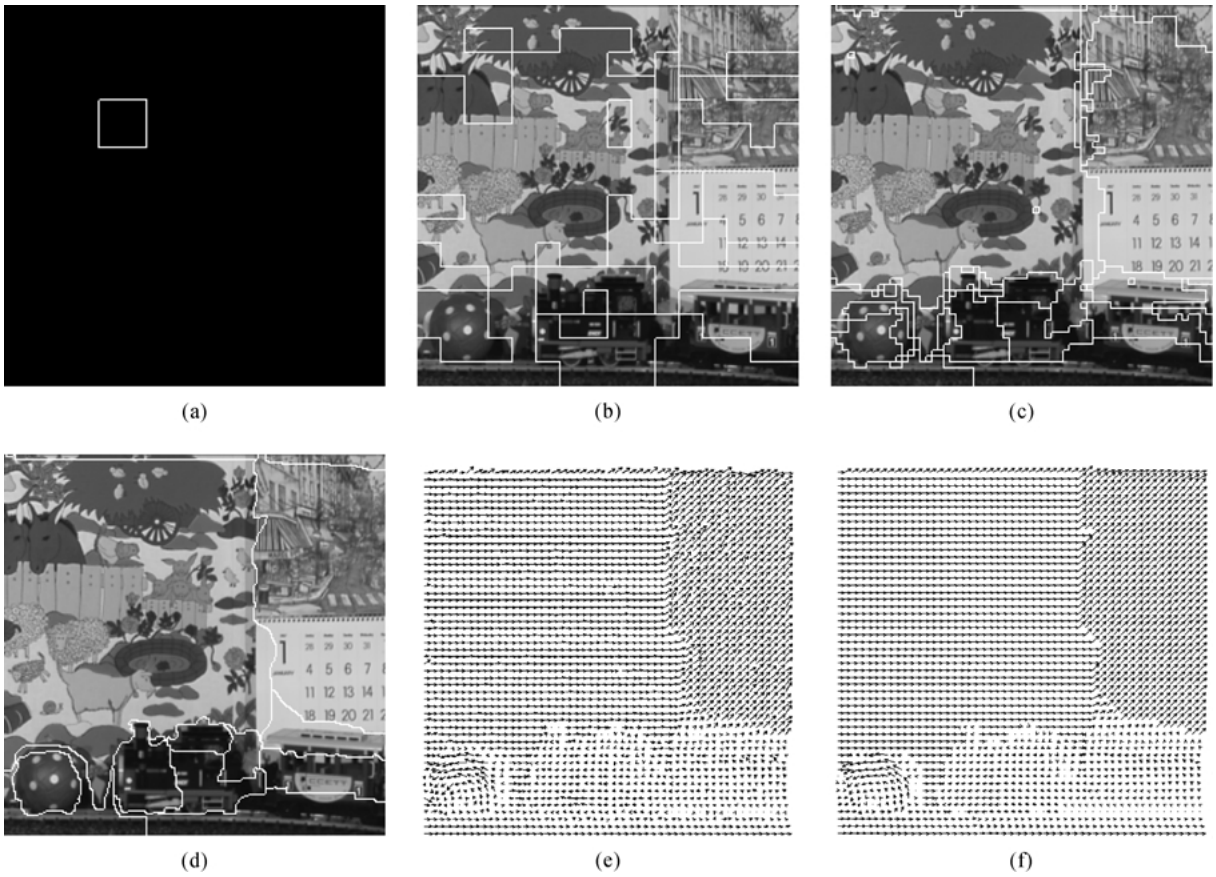


Figure 14. Joint motion estimation-segmentation on *Calendar*: (a) partition initialization at level $\ell = 4$; (b–d) final segmentations at grid levels $\ell = 4, 2, 0$; (e) final dense field estimate; (f) parametric field estimate (cpu time $\sim$8 mn).

5. Conclusion

In this paper, we have presented a comprehensive energy-based framework for the incremental estimation and segmentation of apparent motion fields. Using robust cost functions, a dense discontinuity-preserving motion estimation technique has first been introduced, and a special care has been dedicated to its algorithmic implementation: a hierarchical constrained minimization framework is proposed which allows to mix different increment parameterizations with respect to a regular or an adaptive partitioning of the image. The ability of the resulting method to recover intricate non-rigid motions has been especially demonstrated on sequences involving moving fluids. For situations where a motion-based segmentation of the sequence makes sense and is of interest, the previous model has been extended to simultaneously handle both tasks. A dense estimation as well as a parametric representation of the same motion field are thus jointly recovered in an alternate and cooperative way. Of particular interest here, we propose a simple mechanism of interaction between a dense discontinuity-preserving estimation process and a segmentation process, through the auxiliary variables that appear in the half-quadratic formulation of robust cost functions. We believe that this mechanism could be used elsewhere (e.g., simultaneous restoration-segmentation of still images), and could probably be considered (and theoretically studied) from the continuous standpoint of anisotropic diffusion.

The dense motion estimation technique we propose constitutes a generic tool whose flexibility allows

the design of "specialized" versions dedicated to specific motion estimation (or more generally registration) problems. It has been, for example, adapted to the problems of 3D brain image registration Hellier et al. (2001) and of dense stereo matching (Oisel et al., 2000). Based on the promising results it already produced on challenging fluid sequences (as reported in Section 4), we have also started to investigate the incorporation of new modeling ingredients related to this very specific kind of dynamical contents (Corpetti et al., 2000; Mémin and Pérez, 1999).

Concerning the joint estimation-segmentation technique, further research directions include the design of more sophisticated or complete interaction mechanisms (e.g., to take into account photometric discontinuities as a useful cue, or to handle explicitly the problem of occlusions at borders of motion regions).

Appendix A

Gauss-Seidel Iteration for Dense Motion Estimation under Parametric Constraint

For the sake of concision, we shall denote $\nabla \tilde{f}(s) \triangleq \nabla f(s+\boldsymbol{w}_s^\ell, t+1)$ the spatial gradient in the second image, displaced according to $\boldsymbol{w}^\ell$, and $\tilde{f}_t(s) = f_t(s, \boldsymbol{w}_s^\ell)$ the displaced frame difference. S is partitioned according to $\mathcal{B}^\ell = \{\mathcal{B}_1^\ell \dots \mathcal{B}_{N_\ell}^\ell\}$. Let $\mathcal{B}_n^\ell$ be the current block in the iterative visits performed by the Gauss-Seidel solver. One has simply to minimize $\mathcal{H}^\ell$ with respect to $\boldsymbol{\theta}_n^\ell$, the total field outside $\mathcal{B}_n^\ell$ being frozen. The fraction of energy actually concerned is:

$$\begin{aligned}\mathcal{H}_n^\ell\big(\boldsymbol{\theta}_n^\ell, \delta, \beta\big) \triangleq\ & \tau_1 \sum_{s\in\mathcal{B}_n^\ell} \delta_s \big[\nabla \tilde{f}(s)^T P_n(s)\boldsymbol{\theta}_n^\ell + \tilde{f}_t(s)\big]^2 \\ & + \alpha\tau_2 \sum_{\langle s,r\rangle\in\mathcal{C}_{\partial n}^\ell} \beta_{sr} \big\|\boldsymbol{w}_s^\ell + P_n(s)\boldsymbol{\theta}_n^\ell - \boldsymbol{w}_r\big\|^2 \\ & + \alpha\tau_2 \sum_{\langle s,r\rangle\in\mathcal{C}_n^\ell} \beta_{sr} \big\|\big(\boldsymbol{w}_s^\ell + P_n(s)\boldsymbol{\theta}_n^\ell\big) \\ & - \big(\boldsymbol{w}_r^\ell + P_n(r)\boldsymbol{\theta}_n^\ell\big)\big\|^2, \end{aligned} \tag{39}$$

where $\mathcal{C}_{\partial n}^\ell \triangleq \cup_m \mathcal{C}_{nm}^\ell$ is the set of pairs in $\mathcal{C}$ straddling the border of $\mathcal{B}_n^\ell$. The increment field in the neighborhood of $\mathcal{B}_n^\ell$ is a mix of various parameterizations relative to the different parts of the (possibly irregular) grid S^ℓ. However, the only thing of actual interest when updating $\boldsymbol{\theta}_n^\ell$ is the total field $\boldsymbol{w}_r \triangleq \boldsymbol{w}_r^\ell + P_m(r)\boldsymbol{\theta}_m^\ell$ at any location r in any neighboring block $\mathcal{B}_m^\ell$. As a consequence, in the following computations, the neighboring parameterizations do not appear explicitly in the regularization part of the update. Their are simply hidden within the total field on the neighboring patches. Letting the partial derivative of this piece of energy vanish yields:

$$\begin{aligned}&\frac{\partial \mathcal{H}_n^\ell\big(\boldsymbol{\theta}_n^\ell, \beta, \delta\big)}{\partial \boldsymbol{\theta}_n^\ell} \\ &= \tau_1 \sum_{s\in\mathcal{B}_n^\ell} \delta_s P_n(s)^T \nabla\tilde{f}(s)\big[\nabla\tilde{f}(s)^T P_n(s)\boldsymbol{\theta}_n^\ell + \tilde{f}_t(s)\big] \\ &\quad + \alpha\tau_2 \sum_{\langle s,r\rangle\in\mathcal{C}_{\partial n}^\ell} \beta_{sr} P_n(s)^T \big[\boldsymbol{w}_s^\ell + P_n(s)\boldsymbol{\theta}_n^\ell - \boldsymbol{w}_r\big] \\ &\quad + \alpha\tau_2 \sum_{\langle s,r\rangle\in\mathcal{C}_n^\ell} \beta_{sr} (P_n(s) - P_n(r))^T \\ &\qquad \times \big[\boldsymbol{w}_s^\ell - \boldsymbol{w}_r^\ell + (P_n(s) - P_n(r))\boldsymbol{\theta}_n^\ell\big] = 0. \end{aligned} \tag{40}$$

A compact vector formulation of this equation can be achieved by introducing the following matrices and vectors indexed respectively by the pixels of block $\mathcal{B}_n^\ell$, the neighbor pairs inside the block, and those straddling the border of the block:

$$A_n \triangleq \begin{bmatrix} \vdots \\ \nabla\tilde{f}(s)^T P_n(s) \\ \vdots \end{bmatrix}_{s\in\mathcal{B}_n^\ell}, \quad \boldsymbol{F}_n \triangleq \begin{bmatrix} \vdots \\ \tilde{f}_t(s) \\ \vdots \end{bmatrix}_{s\in\mathcal{B}_n^\ell},$$

$$\Delta_n \triangleq \operatorname{diag}(\dots, \delta_s, \dots)_{s\in\mathcal{B}_n^\ell},$$

$$C_n \triangleq \begin{bmatrix} \vdots \\ P_n(s) - P_n(r) \\ \vdots \end{bmatrix}_{\langle s,r\rangle\in\mathcal{C}_n^\ell},$$

$$B_n \triangleq \operatorname{diag}(\dots, \beta_{sr}\mathbb{I}_2, \dots)_{\langle s,r\rangle\in\mathcal{C}_n^\ell},$$

$$C_{\partial n} \triangleq \begin{bmatrix} \vdots \\ P_n(s) \\ \vdots \end{bmatrix}_{\langle s,r\rangle\in\mathcal{C}_{\partial n}^\ell}, \quad \text{and}$$

$$B_{\partial n} \triangleq \operatorname{diag}(\dots, \beta_{sr}\mathbb{I}_2, \dots)_{\langle s,r\rangle\in\mathcal{C}_{\partial m}^\ell},$$

where $\mathbb{I}_2 \triangleq \left[\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right]$, as well as the following block-wise and border-wise averages:

$$\bar{\boldsymbol{\theta}}^{\ell}_{\partial n} \triangleq \frac{1}{b_{\partial n}} \sum_{\langle s,r\rangle \in \mathcal{C}^{\ell}_{\partial n}} \beta_{sr} P_n(s)^T \big(\boldsymbol{w}_r - \boldsymbol{w}^{\ell}_s\big), \quad \text{with } b_{\partial n} \triangleq \sum_{\langle s,r\rangle \in \mathcal{C}^{\ell}_{\partial n}} \beta_{sr}$$

$$\bar{\boldsymbol{\theta}}^{\ell}_{n} \triangleq \frac{1}{b_{n}} \sum_{\langle s,r\rangle \in \mathcal{C}^{\ell}_{n}} \beta_{sr} (P_n(s) - P_n(r))^T \big(\boldsymbol{w}^{\ell}_r - \boldsymbol{w}^{\ell}_s\big), \quad \text{with } b_{n} \triangleq \sum_{\langle s,r\rangle \in \mathcal{C}^{\ell}_{n}} \beta_{sr}.$$

Linear Eq. (40) then reads:

$$\big[\tau_1 A_n^T \Delta_n A_n + \alpha\tau_2 C^T_{\partial n} B_{\partial n} C_{\partial n} + \alpha\tau_3 C_n^T B_n C_n\big]\boldsymbol{\theta}^{\ell}_n = -\tau_1 A_n^T \Delta_n \boldsymbol{F}_n + \alpha\tau_2 b_{\partial n} \bar{\boldsymbol{\theta}}^{\ell}_{\partial n} + \alpha\tau_3 b_n \bar{\boldsymbol{\theta}}^{\ell}_{n}. \quad (41)$$

The direct resolution of this linear system provides the updated value of parameter vector $\boldsymbol{\theta}^{\ell}_n$. In this equation, matrices A_n, C_n and $C_{\partial n}$, and vectors $\bar{\boldsymbol{\theta}}^{\ell}_n$ and $\bar{\boldsymbol{\theta}}^{\ell}_{\partial n}$ depend on the type of parameterization associated with block $\mathcal{B}^{\ell}_n$. Let us give their expressions (when simplified forms are available) for the two different parameterizations.

For *constant model*, $P_n \equiv \mathbb{I}_2$, yielding $A_n^T = [\cdots \boldsymbol{\nabla}\tilde{f}(s) \cdots]_{s\in\mathcal{B}^{\ell}_n}$, $C^T_{\partial n} = [\cdots \mathbb{I}_2 \cdots]$, $C_n = 0$, $\bar{\boldsymbol{\theta}}^{\ell}_n = 0$, and $\bar{\boldsymbol{\theta}}^{\ell}_{\partial n} = \frac{1}{b_{\partial n}} \sum_{\langle s,r\rangle \in \mathcal{C}^{\ell}_{\partial n}} \beta_{sr}(\boldsymbol{w}_r - \boldsymbol{w}^{\ell}_s)$. Equation (41) simplifies as follows:

$$\begin{aligned}
(41) &\Leftrightarrow \Big(\tau_1 A_n^T \Delta_n A_n + \alpha\tau_2 b_{\partial n}\mathbb{I}_2\Big)\boldsymbol{\theta}^{\ell}_n \\
&= -\tau_1 A_n^T \Delta_n \boldsymbol{F}_n + \alpha\tau_2 b_{\partial n}\bar{\boldsymbol{\theta}}^{\ell}_{\partial n} \\
&\Leftrightarrow \Big(\frac{1}{\gamma} A_n^T \Delta_n A_n + \mathbb{I}_2\Big)\boldsymbol{\theta}^{\ell}_n \\
&= \bar{\boldsymbol{\theta}}^{\ell}_{\partial n} - \frac{1}{\gamma} A_n^T \Delta_n \boldsymbol{F}_n, \quad \text{with } \gamma \triangleq \frac{\alpha\tau_2 b_{\partial n}}{\tau_1} \\
&\Leftrightarrow \boldsymbol{\theta}^{\ell}_n = \bar{\boldsymbol{\theta}}^{\ell}_{\partial n} \\
&\quad - \frac{\gamma A_n^T \Delta_n \Big(A_n \bar{\boldsymbol{\theta}}^{\ell}_{\partial n} + \boldsymbol{F}_n\Big) + \det\mathbb{A}_n \bar{\boldsymbol{\theta}}^{\ell}_{\partial n} + \mathrm{com}\mathbb{A}_n A_n^T \Delta_n \boldsymbol{F}_n}{\gamma(\gamma + \mathrm{trace}\mathbb{A}_n) + \det\mathbb{A}_n},
\end{aligned} \quad (42)$$

with $\mathbb{A}_n \triangleq A_n^T \Delta_n A_n$.

For *affine model*, $P_n(s) = \mathbb{I}_2 \otimes \boldsymbol{e}(s)^T$, with $\boldsymbol{e}(s)^T \triangleq [1\ x_s\ y_s]$, yielding the following expressions for the matrices and vectors involved in Eq. (41):

$$A_n^T \Delta_n A_n = \sum_{s\in\mathcal{B}^{\ell}_n} \delta_s (\boldsymbol{\nabla}\tilde{f}(s)\boldsymbol{\nabla}\tilde{f}(s)^T) \otimes (\boldsymbol{e}(s)\boldsymbol{e}(s)^T)$$

$$C^T_{\partial n} B_{\partial n} C_{\partial n} = \mathbb{I}_2 \otimes \sum_{\langle s,r\rangle \in \mathcal{C}^{\ell}_{\partial n}} \beta_{sr} \boldsymbol{e}(s)\boldsymbol{e}(s)^T$$

$$C_n^T B_n C_n = \mathbb{I}_2 \otimes \sum_{\langle s,r\rangle \in \mathcal{C}^{\ell}_{n}} \beta_{sr} (\boldsymbol{e}(s) - \boldsymbol{e}(r))(\boldsymbol{e}(s) - \boldsymbol{e}(r))^T = \mathbb{I}_2 \otimes \mathrm{diag}\left(0, \sum_{\langle s,r\rangle \in \mathcal{C}^{\ell}_{n}(\updownarrow)} \beta_{sr}, \sum_{\langle s,r\rangle \in \mathcal{C}^{\ell}_{n}(\leftrightarrow)} \beta_{sr}\right)$$

$$A_n^T \Delta_n \boldsymbol{F}_n = \sum_{s\in\mathcal{B}^{\ell}_n} \delta_s f_t(s) \boldsymbol{\nabla}\tilde{f}(s) \otimes \boldsymbol{e}(s)$$

$$b_{\partial n}\bar{\boldsymbol{\theta}}^{\ell}_{\partial n} = \sum_{\langle s,r\rangle \in \mathcal{C}^{\ell}_{\partial n}} \beta_{sr}\big(\boldsymbol{w}_r - \boldsymbol{w}^{\ell}_s\big) \otimes \boldsymbol{e}(s)$$

$$b_{n}\bar{\boldsymbol{\theta}}^{\ell}_{n} = \sum_{\langle s,r\rangle \in \mathcal{C}^{\ell}_{n}} \beta_{sr}\big(\boldsymbol{w}^{\ell}_r - \boldsymbol{w}^{\ell}_s\big) \otimes (\boldsymbol{e}(s) - \boldsymbol{e}(r)) = \sum_{\langle s,r\rangle \in \mathcal{C}^{\ell}_{n}(\updownarrow)} \beta_{sr}\big(\boldsymbol{w}^{\ell}_r - \boldsymbol{w}^{\ell}_s\big) \otimes [0\ 1\ 0]^T + \sum_{\langle s,r\rangle \in \mathcal{C}^{\ell}_{n}(\leftrightarrow)} \beta_{sr}\big(\boldsymbol{w}^{\ell}_r - \boldsymbol{w}^{\ell}_s\big) \otimes [0\ 0\ 1]^T.$$

where $\mathcal{C}^{\ell}_n(\updownarrow)$ (resp. $\mathcal{C}^{\ell}_n(\leftrightarrow)$) contains pixel pairs of $\mathcal{C}^{\ell}_n$ lying along the x-direction (resp. y-direction).

Appendix B

Constrained Estimation-Segmentation at Grid Level ℓ

Data term: Using the same block-wise notations as in Appendix A, it is easy to get the following compact expression:

$$\mathcal{H}^{\ell}_1(\mathsf{d}\boldsymbol{w}^{\ell}, \delta) = \sum_{n\in S^{\ell}} \Bigg[\tau_1\big(A_n \mathsf{d}\boldsymbol{w}^{\ell}_n + \boldsymbol{F}_n\big)^T \times \Delta_n\big(A_n \mathsf{d}\boldsymbol{w}^{\ell}_n + \boldsymbol{F}_n\big) + \sum_{s\in\mathcal{B}^{\ell}_n} \psi_1(\delta_s)\Bigg]. \quad (43)$$

For each site of S^{ℓ}, one gets a sort of block-wise optical flow expression involving aggregated observations.

Smoothing term: Considering the piece-wise constant constraint on the increment field, the prior energy can be written as Mémin and Pérez (1998a):

$$\mathcal{H}^{\ell}_2(\mathsf{d}\boldsymbol{w}^{\ell}, \beta) = \mathcal{H}_2(\mathbf{0}, \beta) + \tau_2 \sum_{\langle n,m\rangle} \Big[\beta_{nm} \big\|\mathsf{d}\boldsymbol{w}^{\ell}_n - \mathsf{d}\boldsymbol{w}^{\ell}_m\big\|^2 + 2\big(\mathsf{d}\boldsymbol{w}^{\ell}_n - \mathsf{d}\boldsymbol{w}^{\ell}_m\big)^T \overline{\Delta\boldsymbol{w}}^{\ell}_{nm}\Big], \quad (44)$$

with $\beta_{nm} \triangleq \sum_{\langle s,r\rangle \in \mathcal{C}^{\ell}_{nm}} \beta_{sr}$ and $\overline{\Delta\boldsymbol{w}}^{\ell}_{nm} \triangleq \sum_{\langle s,r\rangle \in \mathcal{C}^{\ell}_{nm}} \beta_{sr}(\boldsymbol{w}^{\ell}_s - \boldsymbol{w}^{\ell}_r)$.

Parametric likeness term: $\mathcal{R}^\ell \in \Lambda^\ell$ being the current partition of S^ℓ with associated parameter vector $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_i)$, we denote $\mathcal{R} = \Psi(\mathcal{R}^\ell)$ the associated constrained partition of S and $\partial\mathcal{R}_{ij}$ the pieces of frontiers between adjacent regions $\mathcal{R}_i$ and $\mathcal{R}_j$ of $\mathcal{R}$. We have:

$$\begin{aligned}\mathcal{H}_3^\ell(\mathsf{d}\boldsymbol{w}^\ell, \mathcal{R}^\ell, \boldsymbol{\varphi}, \beta, \eta) &= \mu_1 \sum_{\langle i,j\rangle} \frac{1}{|\partial\mathcal{R}_{ij}|} \sum_{\langle s,r\rangle \in \partial\mathcal{R}_{ij}} \beta_{sr} + \mu_2 \sum_i \sum_{n\in\mathcal{R}_i^\ell} \sum_{s\in\mathcal{B}_n^\ell} \\ &\quad \times \left[\tau_3 \eta_s \left\| \boldsymbol{w}_s^\ell + \mathsf{d}\boldsymbol{w}_n^\ell - P(s)\boldsymbol{\varphi}_i \right\|^2 + \psi_3(\eta_s)\right], \quad (45)\end{aligned}$$

which, like previously, reduces to:

$$\begin{aligned}\mathcal{H}_3(\mathsf{d}\boldsymbol{w}^\ell, \mathcal{R}^\ell, \boldsymbol{\varphi}, \beta, \eta) &= \mathcal{H}_3^\ell(\mathbf{0}, \mathcal{R}^\ell, \boldsymbol{\varphi}, \beta, \eta) + \mu_2 \sum_i \sum_{n\in\mathcal{R}_i} \left[z_n \tau_3 \left\| \mathsf{d}\boldsymbol{w}_n^\ell \right\|^2 \right. \\ &\quad \left. + 2\left(\mathsf{d}\boldsymbol{w}_n^\ell\right)^T \overline{\Delta \boldsymbol{w}}_n^\ell(\boldsymbol{\varphi}_i)\right], \quad (46)\end{aligned}$$

where $z_n \triangleq \sum_{s\in\mathcal{B}_n^\ell} \eta_s$ and $\overline{\Delta \boldsymbol{w}}_n^\ell(\boldsymbol{\varphi}_i) \triangleq \sum_{s\in\mathcal{B}_n^\ell} \eta_s(\boldsymbol{w}_s^\ell - P(s)\boldsymbol{\varphi}_i)$.

Segmentation a priori term: The reduced expression of this term is:

$$\mathcal{H}_4(\mathcal{R}^\ell) = \mathcal{H}_4^\ell(\boldsymbol{\Psi}^\ell(\mathcal{R}^\ell)) = \lambda 2^\ell |\partial\mathcal{R}^\ell|. \quad (47)$$

Gauss-Seidel iteration w.r.t. $\mathsf{d}\boldsymbol{w}_n^\ell$: Setting to zero the derivative of $\mathcal{H}^\ell$ w.r.t. to $\mathsf{d}\boldsymbol{w}_n^\ell$, one gets the same update equation as in (42) (with $\mathsf{d}\boldsymbol{w}_n^\ell = \boldsymbol{\theta}_n^\ell$, for it is the constant model), but with slightly modified definitions:

$$\begin{aligned}\bar{\boldsymbol{\theta}}_{\partial n}^\ell &\triangleq \frac{\sum_{m:\mathcal{C}_{mn}^\ell \neq \emptyset} \left(\beta_{nm}\mathsf{d}\boldsymbol{w}_m^\ell - \overline{\Delta \boldsymbol{w}}_{nm}^\ell\right) - \frac{\mu_2\tau_3}{\alpha\tau_1}\overline{\Delta \boldsymbol{w}}_n^\ell(\boldsymbol{\varphi}_i)}{b_{\partial n} + \frac{\mu_2\tau_3}{\alpha\tau_1} z_n}, \\ \gamma &\triangleq \frac{\alpha\tau_2 b_{\partial n} + \mu_2\tau_3 z_n}{\tau_1},\end{aligned} \quad (48)$$

where now appears the parametric field of the region $\mathcal{R}_i$ to whom block n belongs. For $\eta \equiv 0$ (which implies $z_n = 0$ and $\overline{\Delta \boldsymbol{w}}_n^\ell(\boldsymbol{\varphi}_i) \equiv 0$), the expression coincides with (42), as expected, since this amounts to removing parametric goodness-of-fit energy term.

Notes

1. Function ψ is defined as $\psi(z) \triangleq \phi \circ \phi'^{-1}(\tau z) - \tau z \phi'^{-1}(\tau z)$ (Charbonnier et al., 1997; Geman and Reynolds, 1992; Mémin and Pérez, 1998a). It is strictly decreasing since $\psi'(z) = -\tau\phi'^{-1}(\tau z) < 0$.
2. In case $\mathcal{B}$ is a regular partition into square patches, the adjacency graph is the N-site rectangular lattice with same neighborhood system as the original lattice.
3. By reference to the standard multigrid techniques from numerical analysis (Hackbusch, 1985) to which our hierarchical minimization scheme is related we will say that ℓ indexes *grid levels*. These grid levels are not to be confused with the *resolution levels*: at each level of resolution a complete sequence of grid levels from $\ell = L$ to 0 is deployed to conduct the minimization.
4. A natural way of building this hierarchy of parametric representations is to consider nested partitions where $\mathcal{B}^\ell$ is made up from the subdivision of elements of $\mathcal{B}^{\ell+1}$. This nested structure is easily obtained with regular subdivision schemes (based on square or triangle tiling). It is more difficult to design irregular subdivision strategies. In Section 4, we shall introduce an adaptive way to build square-based nested partitions.
5. When the constrained subsets are nested, i.e., $\Omega^{\ell+1} \subset \Omega^\ell$, the succession of minimizations can be conducted in a slightly different way: the final estimate at a given level is not directly integrated in the main field to be refined at the next level, but simply used as an initialization for the iterative minimization process. More precisely, all $\boldsymbol{w}^\ell$ fields in (13) are the same, equal to some field $\boldsymbol{w}$, while $\hat{\boldsymbol{\theta}}^{\ell+1}$ is now used to define the initial increment configuration at level ℓ through $(\Phi^\ell)^{-1} \circ \Phi^{\ell+1}(\hat{\boldsymbol{\theta}}^{\ell+1})$, which makes sense since $\mathrm{Im}\Phi^{\ell+1} \subset \mathrm{Im}\Phi^\ell$. In this version, described in Mémin and Pérez (1998a), the spatio-temporal luminance derivatives remain the same, i.e., computed with respect to $f(s,t)$ and $f(s+\boldsymbol{w}_s, t+1)$. In other terms, a single linearization of the brightness constancy assumption is considered (for a given resolution level), and the coarse-to-fine minimization turns out to be a standard *multigrid* scheme (Hackbusch, 1985).
6. We still consider in the coming developments that an arbitrary resolution level of the multiresolution setting is concerned.
7. As mentioned in Section 2.3, this hierarchical Gauss-Newton minimization can be replaced by a classic *multigrid* minimization when the constrained configuration subsets are nested, which is the case here ($\Omega^{\ell+1} \times \Upsilon^{\ell+1} \subset \Omega^\ell \times \Upsilon^\ell$): in this variant the final estimate at level $\ell+1$ is projected at level ℓ through $[\Phi^\ell]^{-1} \circ \Phi^{\ell+1}$ (resp. $[\Psi^\ell]^{-1} \circ \Psi^{\ell+1}$) and used as an initial configuration at that level ℓ.
8. The sequence and the estimated vector fields can be seen at `http://www.irisa.fr/vista/Demos/Demos.english.html`.

References

Adiv, G. 1985. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Trans. Pattern Anal. Machine Intell.*, 7:384–401.

Ayer, S. and Sawhney, H.S. 1995. Layered representation of motion video using robust maximum-likelihood estimation of mixture models and Mdl encoding. In *Proc. Int. Conf. Computer Vision*, pp. 777–784.

Bab-Hadashiar, A. and Sutter, D. 1998. Robust optic flow computation. *Int. J. Computer Vision*, 29(1):59–77.

Barron, J., Fleet, D., and Beauchemin, S. 1994. Performance of optical flow techniques. *Int. J. Computer Vision*, 12(1):43–77.

Bergen, J., Anandan, P., Hanna, K., and Hingorani, R. 1992. Hierarchical model-based motion estimation. In *Proc. Europ. Conf. Computer Vision*, G. Sandini (Ed.), vol. 558 of LNCS, pp. 237–252, Springer-Verlag: Berlin.

Besag, J. 1986. On the statistical analysis of dirty pictures. *J. Royal Statist. Soc.*, 48B(3):259–302.

Black, M. 1994. Recursive non-linear estimation of discontinuous flow fields. In *Proc. Europ. Conf. Computer Vision*, Stockholm, Sweden, pp. 138–145.

Black, M. and Anandan, P. 1996. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104.

Black, M. and Jepson, P. 1996. Estimating optical flow in segmented images using variable-order parametric models with local deformations. *IEEE Trans. Pattern Anal. Machine Intell.*, 18(10):972–986.

Black, M. and Rangarajan, A. 1996. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *Int. J. Computer Vision*, 19(1):75–104.

Bouthemy, P. and Francois, E. 1993. Motion segmentation and qualitative dynamic scene analysis from an image sequence. *Int. J. Computer Vision*, 10(2):157–182.

Chang, M.M., Tekalp, A.M., and Sezan, M.I. 1997. Simultaneous motion estimation and segmentation. *IEEE Trans. Image Processing*, 6(9):1326–1333.

Charbonnier, P., Blanc-Féraud, L., Aubert, G., and Barlaud, M. 1997. Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Image Processing*, 6(2):298–311.

Cohen, I. and Herlin, I. 1999. Non uniform multiresolution method for optical flow and phase portrait models: Environmental applications. *Int. J. Computer Vision*, 33(1):29–49.

Corpetti, T., Mémin, E., and Pérez, P. 2000. Dense fluid flow estimation. IRISA, Technical Report No. 1352.

Delanay, A. and Bresler, Y. 1998. Globally convergent edge-preserving regularized reconstruction: An application to limited-angle tomography. *IEEE Trans. Image Processing*, 7(2):204–221.

Deriche, R., Kornprobst, P., and Aubert, G. 1995. Optical flow estimation while preserving its discontinuities: A variational approach. In *Proc. Asian Conf. Computer Vision*, Singapore, vol. 1, pp. 290–295.

Enkelmann, W. 1988. Investigation of multigrid algorithms for the estimation of optical flow fields in image sequences. *Comp. Vision Graph. and Image Proces.*, 43:150–177.

Geman, D. and Reynolds, G. 1992. Constrained restoration and the recovery of discontinuities. *IEEE Trans. Pattern Anal. Machine Intell.*, 14(3):367–383.

Hackbusch, W. 1985. *Multi-Grid Methods and Applications*. Springer-Verlag: Berlin.

Hellier, P., Barillot, C., Mémin, E., and Pérez, P. (2001). Hierarchical estimation of a dense deformation field for 3D robust registration. *IEEE Transaction on Medical Imaging*, 20(5):388–402.

Horn, B. and Schunck, B. 1981. Determining optical flow. *Artificial Intelligence*, 17:185–203.

Huang, Y., Palaniappan, K., Zhuand, X., and Cavanaugh, J. 1995. Optic flow field segmentation and motion estimation using a robust genetic partitioning algorithm. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(12):1177–1190.

Ju, X., Black, M.J., and Jepson, A.D. 1996. Skin and bones: Multi-layer, locally affine, optical flow and regularization with transparency. In *Proc. Conf. Comp. Vision Pattern Rec.*, pp. 307–314.

Kornprobst, P., Deriche, R., and Aubert, G. 1999. Image sequence analysis via partial differential equations. *Journal of Mathematical Imaging and Vision*, 11(1):5–26.

Lai, S. and Vemuri, B. 1998. Reliable and efficient computation of optical flow. *Int. J. Computer Vision*, 29(2):87–105.

Leclerc, Y. 1989. Constructing simple stable descriptions for image partitioning. *Int. J. Computer Vision*, 3:73–102.

Mémin, E. and Pérez, P. 1998a. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Trans. Image Processing*, 7(5):703–719.

Mémin, E. and Pérez, P. 1998b. A multigrid approach for hierarchical motion estimation. In *Proc. Int. Conf. Computer Vision*, Bombay, India, pp. 933–938.

Mémin, E. and Pérez, P. 1999. Fluid motion recovery by coupling dense and parametric motion fields. In *Int. Conf. on Computer, ICCV'99*, pp. 620–625.

Mitiche, A. and Bouthemy, P. 1996. Computation of image motion: A synopsis of current problems and methods. *Int. Journ. of Comp. Vis.*, 19(1):29–55.

Mumford, D. and Shah, J. 1989. Optimal approximation by piecewise smooth functions and associated variational problems. *Comm. Pure and Appl. Math.*, 42:577–685.

Murray, D.W. and Buxton, H. 1987. Scene segmentation from visual motion using global optimization. *IEEE Trans. Pattern Anal. Machine Intell.*, 9(2):220–228.

Musse, O., Heitz, F., and Armspach, J.P. 1999. 3D deformable image matching using multiscale minimization of global energy functions. In *Proc. Conf. Comp. Vision Pattern Rec.*, Fort Collins, Colorado, vol. 2, pp. 478–485,

Nesi, P. 1993. Variational approach to optical flow estimation managing discontinuities. *Image and Vision Computing*, 11(7):419–439.

Oisel, L., Mémin, E., and Morin, L. 2000. Geometric driven optical flow estimation and segmentation for 3d reconstruction. In *Proc. Europ. Conf. Computer Vision*, Dublin, Irland, and D. Vernon, (Ed.), vol. II, pp. 847–863. Springer Notes in Computer Science vol. 1843.

Samson, C., Blanc-Ferraud, L., Aubert, G., and Zerubia, J. 1999. Simultaneous image classification and restoration using a variational approach. In *Proc. Conf. Comp. Vision Pattern Rec.*, Fort Collins, Colorado, vol. 2, pp. 618–623.

Schnorr, C., Sprengel, R., and Neumann, B. 1996. A variational approach to the design of early vision algorithms. *Computing Suppl.*, vol. 11, pp. 149–165.

Schunck, B.G. 1986. The image flow constraint equation. *Comput. Vision, Graphics, Image Proc.*, 35:20–46.

Stiller, C. 1997. Object-based estimation of dense motion fields. *IEEE Trans. Image Processing*, 6(2):234–250.

Szeliski, R. and Coughlan, J. 1994 Hierarchical spline-based image registration. In *Proc. Conf. Comp. Vision Pattern Rec.*, Seattle, Washington, pp. 194–201.

Szeliski, R. and Shum, H.-Y. 1996. Motion estimation with quadtree splines. *IEEE Trans. Pattern Anal. Machine Intell.*, 18(12):1199–1210.

Thisted, R.A. 1988. *Elements of Statistical Computing*. Chapman and Hall: London.

Viéville, T. and Faugeras, O. 1992. Robust and fast computation of unbiased intensity derivatives in images. In *Proc. Europ. Conf. Computer Vision*, S. Margherita Ligure, pp. 203–212.

Yezzi, A., Tsai, A., and Willsky, A. 1999. A statistical approach to image segmentation for bimodal and trimodal imagery. In *Proc. Int. Conf. Computer Vision*, Corfu, Greece.

Zhu, S. and Yuille, A. 1996. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Trans. Pattern Anal. Machine Intell.*, 18:884–900.