A Distributed-Parameter Approach to Dynamic Image Motion

Annette Stahl, Paul Ruhnau, Christoph Schnörr

Computer Vision, Graphics, and Pattern Recognition Group
Department of Mathematics and Computer Science;
University of Mannheim, D-68131 Mannheim, Germany

{astahl,ruhnau,schnoerr}@uni-mannheim.de
http://www.cvgpr.uni-mannheim.de

Abstract. The most elementary knowledge available in connection with image motion computation can be informally expressed as "structures do not jump". We present a novel PDE-based representation of image motion exploiting this knowledge. Our distributed-parameter approach takes into account spatial context, unlike Kalman filters applied to point features separately. It performs spatio-temporal regularization in a recursive online fashion, unlike previous variational approaches evaluating entire spatio-temporal image volumes in a batch processing mode. Deviations from the expected velocity distribution generate vector fields that may serve as attentional mechanism for a superordinate processing stage. We briefly speculate about relations of our approach to perceptual phenomena like motion aftereffects.

1 Introduction

1.1 Overview

The most elementary knowledge available in connection with image motion computation can be informally expressed as "structures do not jump". The physics underlying any motion phenomena embodies some inertia leading to smooth velocities. Maybe starting with the work of Heel [11, 12], this knowledge has been exploited in countless applications of computer vision during the last two decades, mainly in terms of the Kalman filter, and more recently through it's modern extensions [9].

A key property of the Kalman filter is that it applies to *lumped systems*, where the system dynamics is summarized by a set of variables which are a function of time alone. As a consequence, when applied to raw pixel intensities for image motion computation, for example, or to depth estimates for surface reconstruction, this leads to the well-known recursive update equations for computing corresponding estimates at each pixel location, *without* any *spatial* context, however. To achieve the latter, *parametric* models, e.g. of surface patches in the scene, have to be assumed, which not only requires far more *specific* knowledge about

the scene, but also necessitates transferring the Kalman formalism into a corresponding parameter space, typically through nonlinear equations relating the directly observable visual measurements to the parameters as new system variables.

Our present work is an attempt to reconsider early work on dynamic motion vision [11], directly at the raw pixel level, in order to attain the following goal:

(i) We want to take into account *spatial context*, leading to a *distributed parameter approach* with states governed by a partial differential equations (PDEs), rather than ordinary differential equations as for lumped systems.
(ii) The approach should exploit the aforementioned elementary knowledge in connection with motion computation. That is, moving structures should exhibit some inertia in it's most rudimentary form: velocities are not expected to change. The computational structure should be *recursive*, enabling online motion computation.

Our ansatz is based on modelling a fictive compressible fluid, using the Burgers equation [6], in combination with a convex regularization approach to image motion computation [17, 19]. We show that the resulting model exhibits, in a distributed-parameter fashion, properties analogous to a dynamic filter for the most elementary polynomial kinematical model of a point feature (cf. [1]): in the absence of image measurements, and if spatial context does not indicate otherwise, velocities do not change.

The approach also admits a control-theoretic interpretation: if image measurements indicate changes of the current velocity distribution, fictive forces modify the system state accordingly. The presence of such forces may serve as an attentional mechanism notifying a higher-level processing stage about unexpected motion events. Finally, it is tempting to point out a potential relevance of our approach for related models of visual perception [16]. However, as we are not experts in this field, we confine ourselves to few speculative remarks.

1.2 Further Related Work

Concerning property (i) stated above, it is well known that *non-local* variational approaches, particularly those exploiting *spatio-temporal* context [20], lead to robust estimates of coherent motion patterns. On the other hand, such approaches are *static* in that an entire spatio-temporal image volume has to be processed at once in batch mode. In contrast, our approach is *dynamic* and leads to recursive computations in online mode, while still performing temporal regularization.

A prominent example for other work exploiting models from fluid mechanics is the work of Christensen on image registration [7] which subsequently instigated a lot of related research (cf. [2] and references therein). The authors also use the momentum equation as prior knowledge. However, while we merely use the material derivative of the velocity (in order to mimic inertia), they consider the opposite special case of very low Reynolds numbers and drop the inertial terms. Furthermore, as we consider relatively small displacements between video frames, we do not distinguish between velocity and displacements.

Other related work concerns the vorticity transport formulation of the *incompressible* Navier-Stokes equation used for image inpainting [3]. There are two conceptual differences to our approach: First, the inpainting approach transports a function of the gray value field while we transport the velocity field itself. Second, we utilize a *compressible* model of fluids, because incompressibility almost never holds for 2D-deformations induced by relative motions to a 3D scene.

An interesting control-approach to image motion estimation has been proposed in [4]. The basic idea is to estimate both an optical flow field u and a rectified image function $\tilde{g}$ satisfying (2) below exactly. Solving for the adjoint equation, however, requires to move backwards in time and does not enable online computation.

Finally, we point out that while our approach performs spatio-temporal filtering and may be regarded as a rudimentary version of a *distributed-parameter* Kalman filter, it falls short of an exact corresponding mathematical extension (cf., e.g. [5]). The advantage, however, is a deterministic and easy-to-parallelize recursive computational architecture.

2 Approach

2.1 The 2D Burgers Equation

Let $u = (u_1, u_2)^\top$, $u = u(x,t)$, $x = (x_1(t), x_2(t))$, denote a two-dimensional vector field. We consider a hyperbolic system of conservation laws in terms of the inviscid Burgers equation

$$\frac{D}{Dt}u = \frac{\partial u}{\partial t} + (u \cdot \nabla)u = 0 , \quad u(x,0) = u_0 \tag{1}$$

This equation has been studied and successfully applied for many decades in aero- and fluid dynamics [6, 21, 15, 10, 13] as a simplified model for turbulence, boundary layer behavior, shock wave formation and mass transport. It appears as convection term in the fundamental equations of fluid mechanics, the Navier-Stokes equations.

As a physical interpretation, u in (1) may be regarded as a vector of conserved (fictive) quantities or states, with corresponding density functions u_1, u_2 as components. The material derivative $\frac{D}{Dt}$ in (1) yields the acceleration of moving particles. Consequently, equation (1) says that velocities of moving particles do not change. The nonlinear term $(u \cdot \nabla)u$ is the inertia term of the transport process described by (1). See figure 1 for an illustration. In practice, we numerically solve (1) for the (short) time-interval $[0, T]$ between two video frames (section 3).

2.2 Variational Model

Let $g(x_1, x_2, t)$ denote the gray value of an image sequence recorded at location $x = (x_1, x_2)^\top$ within some rectangular image domain Ω and time $t \in [0, T]$.
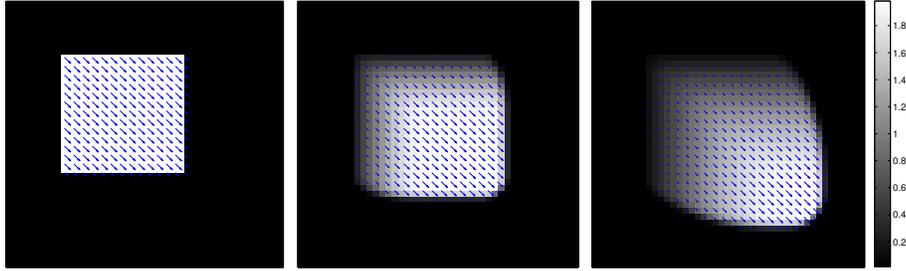
Fig. 1. Transportation of a vector field by equation (1) at times $t = 0, 5, 10$. Gray values visualize vector magnitudes. Fictive particles move along a shock front in the lower right direction. In the absence of any further external information, a region of rarefaction arises due to mass conservation, acting like a short-time memory.

We adopt the common basic assumption underlying most approaches to motion estimation that g is conserved over time:

$$\frac{D}{Dt}g = (u \cdot \nabla)g + \partial_t g = 0 \tag{2}$$

Here, ∇ denotes the spatial gradient, and ∂_t the partial derivative wrt. time. In order to regularize (2), we adopt the variational approach of Nagel [17]:

$$\frac{1}{2}\int_{\Omega} \left\{ \left((u \cdot \nabla)g + \partial_t g\right)^2 + \alpha^2 \left(\nabla u_1^\top D(\nabla g)\nabla u_1 + \nabla u_2^\top D(\nabla g)\nabla u_2\right) \right\} dx \tag{3}$$

where

$$D(\nabla g) = \frac{1}{\|\nabla g\|^2 + 2\lambda^2}\left(\nabla g^\perp \left(\nabla g^\perp\right)^\top + \lambda^2 I\right), \qquad \mathbb{R} \ni \alpha, \lambda > 0 \tag{4}$$

Under weak assumptions, functional (3) is strictly convex and has a unique minimizing vector field u [18]. The regularizing term, introduced by Nagel, extends the basic approach of Horn and Schunck [14] to an anisotropic image-driven diffusion process (cf. [17, 19]) which prevents the smoothing of motion boundaries at gray value edges.

We emphasize that the approach (3) takes only into account *spatial* context and determines a vector field for a *fixed* point in time $t \in [0, T]$. While it has been shown in [20] that a spatio-*temporal* extension of this class of approaches, using spatio-temporal gradients and a corresponding domain of integration $\int_{\Omega \times [0,T]} \ldots dx dt$, improves both robustness and accuracy of motion estimation, it leads to a batch processing mode where all variables of the entire spatio-temporal volume have to be determined as solution of quite a large linear system of equations.

Our present work is an attempt to change this static viewpoint of image motion processing to a *dynamic* one. To this end, we solve equation (1) for the

time interval $[0, T]$ between a subsequent pair of image frames, where u_0 denotes our current motion estimate. As a result, we obtain a transported vector field $u_T := u(x, T)$ which can be regarded as a *predicted* vector field, based on the constant velocity assumption as discussed in section 2.1.

Furthermore, we complement the spatial variational approach (3) with a corresponding penalty term for the purpose of *temporal* regularization:

$$J(u) = \frac{1}{2} \int_\Omega \left\{ \left((u \cdot \nabla)g + \partial_t g \right)^2 + \alpha^2 \left(\nabla u_1^\top D(\nabla g) \nabla u_1 + \nabla u_2^\top D(\nabla g) \nabla u_2 \right) \right. $$
$$\left. + \beta^2 \|u - u_T\|^2 \right\} dx \tag{5}$$

By minimizing this functional, we combine the predicted state in terms of the vector field u_T with current measurements, given by (2), and spatial regularization. The resulting global minimizer $\overline{u}$ defines the initial data u_0 in (1) for the subsequent image pair and time interval.

2.3 Further Interpretation

An important property of our ansatz (1) and (5) is that temporal context is taken into account without too much additional computational costs. Rather, the solutions to (1) and (5) can be recursively computed in an online processing mode. Of course, regarding temporal regularization, we do *not* claim strict mathematical equivalence to the approach [20].

Problem (5) is solved in a standard way (see [18] for details). The unique global minimizer $\overline{u}$ of the quadratic functional (3) is determined by the variational equation

$$a(\overline{u}, v) = b(v), \quad \forall v \tag{6}$$

where the bilinear form $a(\cdot, \cdot)$ and the linear form $b(\cdot)$ comprise all quadratic and linear terms in u of the functional (3), respectively. A physical interpretation regards (6) as stationary (deformation) state of the "elastic" image domain Ω, defined as equilibrium between the internal deformation energy $a(u, u)$ and external forces $b(u)$, according to the principal of virtual work. Adopting this viewpoint, the additional penalty term in (5) adds an additional internal energy term $\|u\|_{L^2(\Omega)}$ regularizing the degenerate data term, and an additional external force term $\langle u_T, u \rangle_{L^2(\Omega)}$ related to the temporal prediction.

Another viewpoint on (1) and (5) is visual perception. Our approach allows to match computed motions against expectations without making *specific* assumptions about the underlying scene (geometry, rigid/non-rigid, etc.). As a consequence, significant deviations from expected motions may be used to focus the attention of a superordinate processing stage. In section 4, we will inspect this aspect in terms of the vector fields

$$f(x) := \left(u - u_T \right)(x) \quad . \tag{7}$$

3 Implementation Details

We summarize in this section details of our implementation.

3.1 Discretisation of the Burgers Equation

The numerical evaluation of the nonlinear equation (1) is the most involved part
of our approach. We use a second-order conservative finite differencing scheme.
Fluxes are numerically computed by solving the full multidimensional equations
at pixel edges. The correct behavior at discontinuities is obtained by using solu-
tions of the appropriate Riemann problem and by applying van Leer limiters to
some of the second-order terms. For more details, we refer to [8].

3.2 Variational Approach

We use piecewise linear finite elements to discretize the variational equation
(6) related to the variational approach (5), and compute the global minimizer
$\overline{u}$ by solving the resulting linear system of equations. We confine ourselves to
specifying the variational equation more explicitly:

$$a\big((u_1, u_2)^\top, (v_1, v_2)^\top\big) = b\big((v_1, v_2)^\top\big), \quad \forall v_1, v_2 \tag{8}$$

where

$$a(\overline{u}, v) = \int_\Omega \Big\{ \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}^\top \nabla g \nabla g^\top \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

$$+ \alpha^2 \Big(\nabla u_1^\top D(\nabla g) \nabla v_1 + \nabla u_2^\top D(\nabla g) \nabla v_2 \Big) + \beta^2 u^\top v \Big\} dx \tag{9}$$

$$b(v) = \int_\Omega \Big\{ -\partial_t g \nabla g^\top v + \beta^2 \, u_T^\top v \Big\} dx \tag{10}$$

3.3 Coarse-to-Fine Approach and Iterative Registration

The accuracy of motion estimation critically depends on the magnitude of image
motion. It is well known that, depending on the spatial image frequency, very
large motions even may cause aliasing along the time frequency axis. As a rem-
edy, we first compute a coarse velocity field by using only low spatial frequency
components and then "undo" the motion, thus roughly stabilizing the position of
the image over time. Next, the higher frequency subbands are used the estimate
motion on the warped sequence. Combining the resulting correction term with
the previously computed estimate gives a refined velocity estimate.

4 Numerical Experiments

In this section, we examine and illustrate various properties of our approach. All experiments were conducted using a computer-generated scene as ground-truth, where a brick is moving with 7 pixels/frame in front of a wall. We marked the true motion of the brick with a large arrow in the figures below.

We addressed the following aspects:

- Uniform motion conforming to the expected motion, that is no acceleration.
- Unexpected motion events like abrupt starting or stopping movements, or changing motion directions.
- Motion computation with temporarily missing image data in connection (un)expected motion events.
- Temporal regularization.

We discuss these aspects in the following sections.

4.1 Inertia and Motion

Figure 2 illustrates the effect of (1) on motion computation. This effect may be conceived as that of a fictive material, furnished with some inertia, which is trying to follow the apparent motion in an image sequence.

The upper-right panel depicts the frame directly after an abrupt change of the object's moving direction. Due to inertia, the moving particles tend to keep their current motion. After a couple of frames (lower-left panel), a new moving front has been established which drags the remaining field in the novel direction.

Note that we do not regard this temporal period of "confusion" as a drawback. Rather, in most cases where observed motions conform to the expectation, the inertia enables the fictive fluid to exert a temporal filtering effect, as experimentally demonstrated below. Furthermore, the message "nothing interesting happens" can be signaled to a superordinate processing stage. On the other hand, in situations as depicted in figure 2, the discrepancy between observed and expected motion may be used to generate a message so as to focus the attention of superordinate visual processing. The next experiment further illustrates this aspect.

4.2 Focus of Attention and Motion Compression

Figure 3 shows two instantaneous situations for a starting and stopping object, respectively. The vector fields depicted are *not* the optical flows but the vector fields f defined in (7) which quantify deviations from the system's expectation. For example, in the left panel, the vector field points into the reverse direction, due to the "negative acceleration" of the object.

Figure 4 shows a section through f as a function of time (frame number). Two vertical dashed lines indicate where the object stopped and started, respectively. For each event, the curve shows a strong signal peak with very short time delay.

In principle, this peak could be used to focus the attention of a superordinate processing stage.

Figure 5 illustrates this aspect for a real image sequence. The turning points of the waving hand lead to periodic peaks of the vector field f. In between, motion can be predicted well, rendering any communication to external modules unnecessary.

4.3 A Computational Model of Motion Aftereffects?

Another interesting observation is that the curve shown in figure 4 decays with a much lower rate. In a rudimentary way, this is reminiscent of perceptual phenomena like motion aftereffects [16]. After adaption of the visual apparatus to some ongoing stimulus, a "negative" fading visual impression arises when the stimulus disappears. Of course, we are well aware that such an interpretation is highly speculative.

4.4 Missing Image Measurements

Let us consider the behavior of the approach when an entire period of an image sequence is masked out. This scenario is relevant in connection with occlusions or technical limitations in the recording process that lead to data loss.

Figure 6 shows the sequence with a uniformly moving brick where 3 frames have been masked out. Because the motion exactly coincides with the expectation, motion prediction is exact as well, and short-time losses of image data is immaterial.

Things are different when the object changes its direction of movement in the absence of image data. The upper-right panel of figure 7 shows how the algorithm assumes a uniform movement of the object. When image data are again available (lower-left panel), the velocity direction is updated. Notice, however, that there are still (erroneous) velocity estimates below the object showing that the algorithm is still "searching" the moving object in this direction as originally expected.

4.5 Temporal Regularization

Figure 8 compares the L^2-error of the flow estimated with our approach and the motion estimate obtained as minimizer of purely spatial variational approach (3) (i.e. omitting the influence of (1) in (5)) over 20 frames of the sequence. Object motion was uniform. The consistently lower error shows that the ability of motion prediction is used in (5) to exert temporal regularization.
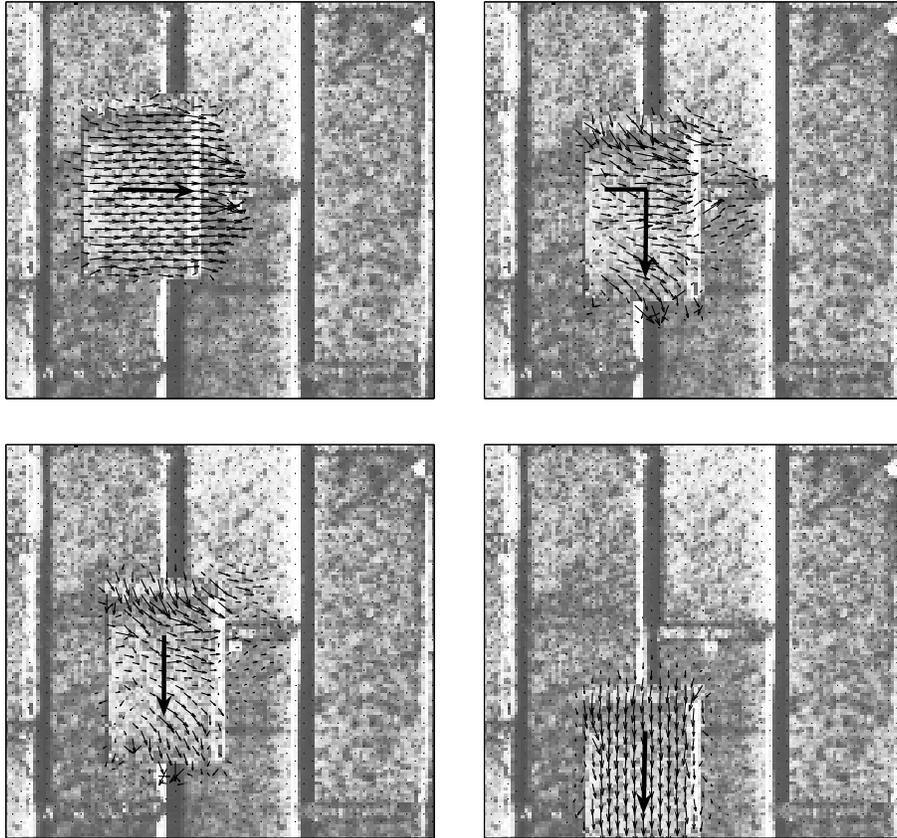
Fig. 2. A moving object which abruptly changes the moving direction, as indicated by the large arrows. The figure illustrates the influence of the inertia of a fictive fluid which is trying to track the apparent image motion. See text for further discussion (parameters: $\alpha = 0.01$, $\beta = 0.005$).

5 Summary and Conclusion

We presented a distributed-parameter approach to dynamic image motion. The approach combines variational motion estimation with motion prediction through a transport process. According to the underlying constant-velocity assumption, the formulation may be regarded as an elementary distributed-parameter version of a low-order polynomial filter commonly used for the kinematics of point features. For motions conforming to the assumption, a temporal regularization effect, computed in a recursive manner, was demonstrated. We also included a brief speculative discussion of our model in view of motion aftereffects.
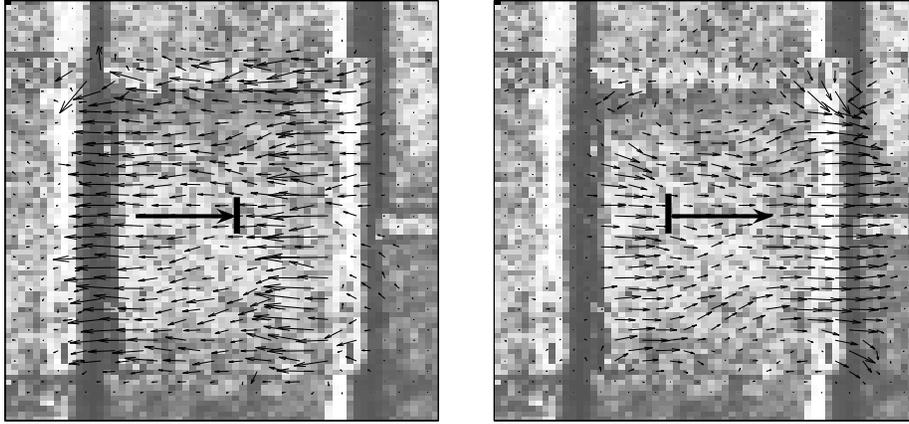
Fig. 3. Left: Vector field f when the object stops. **Right:** Vector field f when the object starts moving. See text for further discussion (parameters: $\alpha = 0.01$, $\beta = 0.001$).
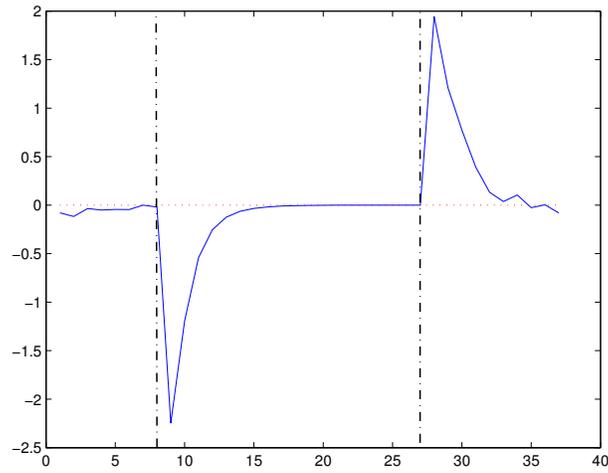

Fig. 4. A section through f (y-axis) over 37 frames (x-axis) corresponding to the image sequence shown in fig. 3. Vertical lines mark the frames where the object started and stopped, respectively. Signal peaks could be used to focuses attention of a superordinate processing level. The fading characteristic is reminiscent of motion aftereffects (parameters: $\alpha = 0.01$, $\beta = 0.001$).

Our further work will address predictions of non-uniform motions and how the approach can be applied in the presence of motion boundaries and occlusions.
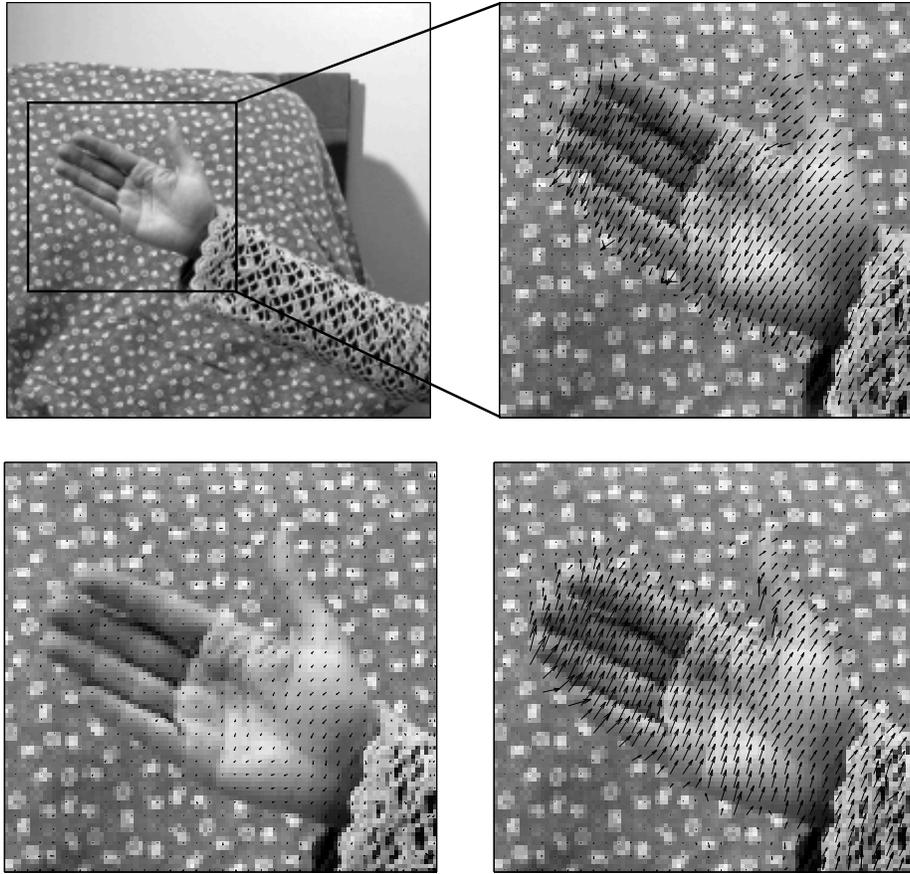
Fig. 5. Top: A waving hand and its the estimated optical flow (right). **Bottom:** At a turning point, the estimated velocity field is zero (left) while the deviation from the expected motion is maximal, causing a "negative" vector field f (right). See text for further discussion (parameters: $\alpha = 0.005$, $\beta = 0.001$).

References

1. Y. Bar-Shalom and T.E. Fortmann. *Tracking and Data Association*. Academic Press, 1988.
2. M.F. Beg, M.I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows. *International Journal of Computer Vision*, 61(2):139–157, 2005.
3. M. Bertalmio, A. Bertozzi, and G. Sapiro. Navier-stokes, fluid dynamics, and image and video inpainting. *Proc. of IEEE-CVPR*, pages 355–362, 2001.
4. A. Borzi, K. Ito, and K. Kunisch. Optimal control formulation for determining optical flow. *SIAM J. Sci. Comput.*, 24(3):818–847, 2002.
5. D. Bosq. *Linear Processes in Function Spaces*, volume 149 of *Lect. Not. Statistics*. Springer, 2000.
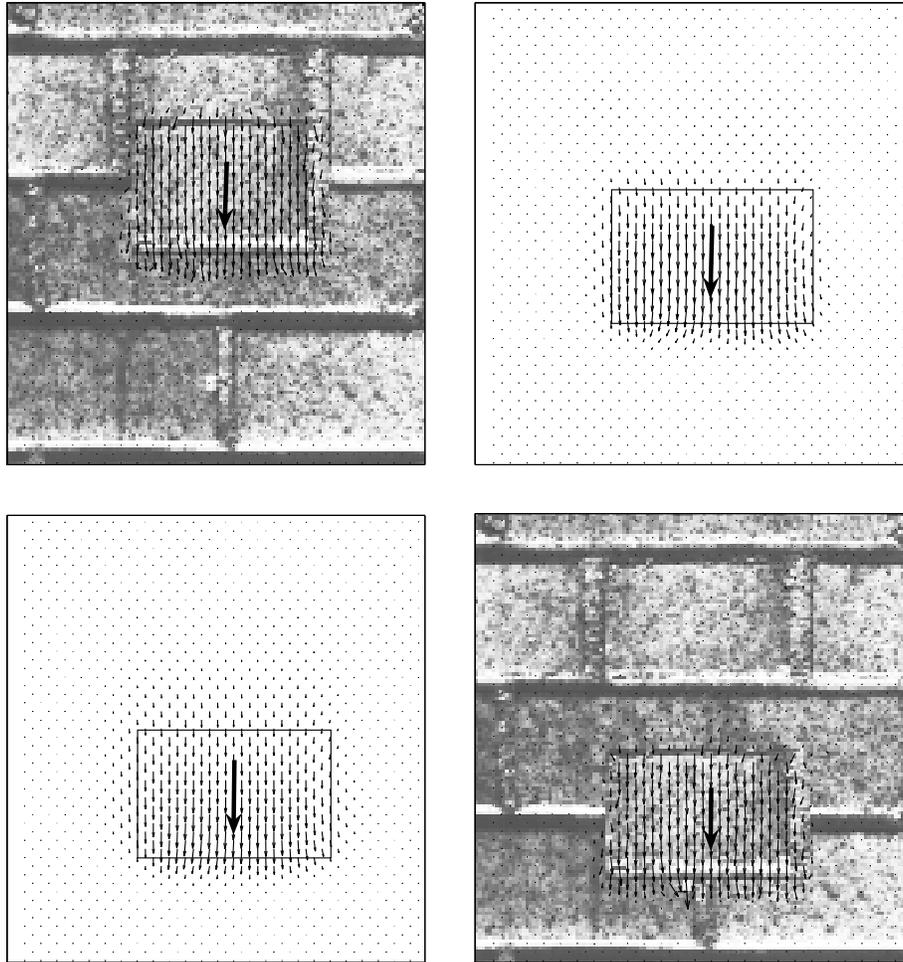
Fig. 6. Flow field estimates for an image sequence with three frames masked out. Because inertia of the fictive material allows for motion prediction, loss of data is immaterial (parameters: $\alpha = 0.01$, $\beta = 0.001$).

6. J.M. Burgers. *A mathematical model illustrating the theory of turbulence*, volume 1. 1948.

7. G.E. Christensen, R.D. Rabbit, and M.I. Miller. Deformable templates using large deformation kinematics. *IEEE Trans. Imag. Proc.*, 5(10):1435–1447, 1996.

8. P. Colella and E.G. Puckett. Modern numerical methods for fluid flow, 1998. http://petroleum.berkeley.edu/UC/PE/modern_numerical_methods_for_flu.htm.

9. A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer Verlag, 2001.

10. C.A.J. Fletcher. *Computational Techniques for Fluid Dynamics (Vol.I)*. Springer-Verlag, 1992.
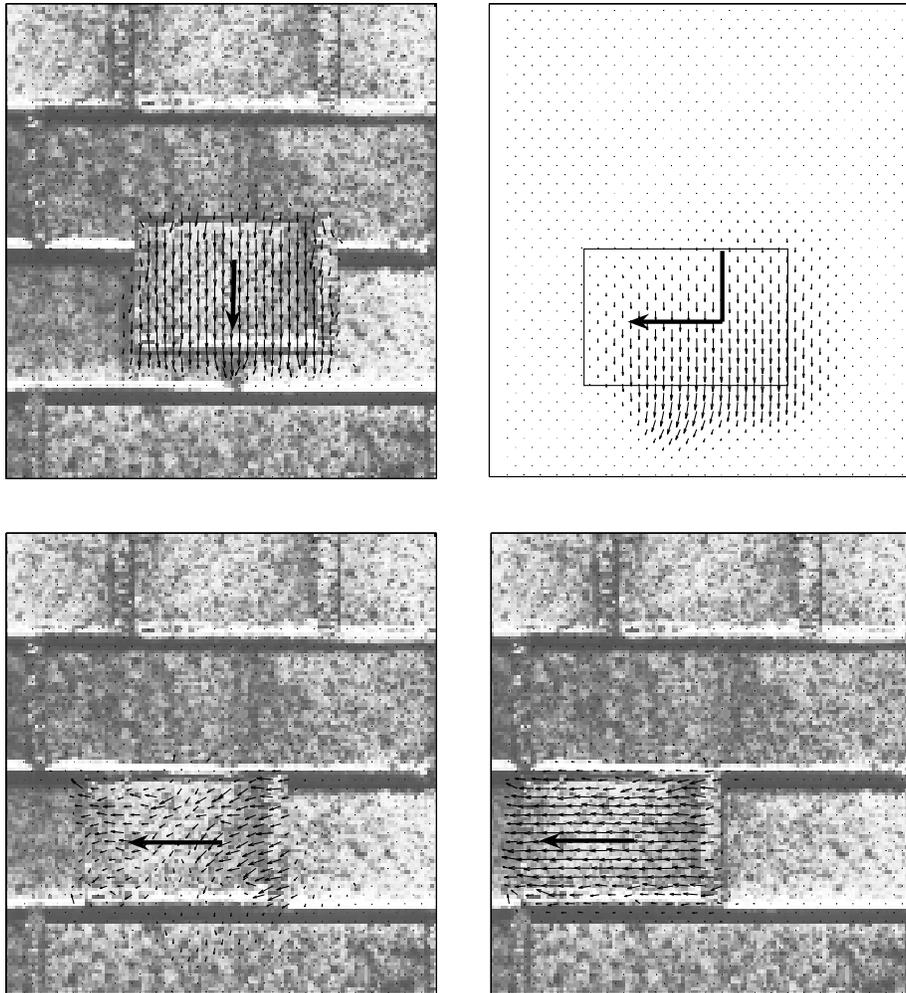
Fig. 7. The opposite situation to figure 6. The object changes its motion during data loss. The flow vectors below the object in the lower-left panel illustrate how the process persists in looking for the moving object in the originally expected direction (parameters: $\alpha = 0.01$, $\beta = 0.001$).

11. J. Heel. Dynamic systems and motion vision. A.I. Memo No. 1037, MIT AI-Lab., 1988.
12. J. Heel. Dynamic motion vision. In *Proc. Image Understanding Workshop*, pages 702–713. Morgan Kaufmann Publ. Inc., Palo Alto, 1989.
13. C. Hirsch. *Numerical Computation of Internal and External Flows (Vol.I+II)*. John Wiley & Sons, 2000.
14. B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.
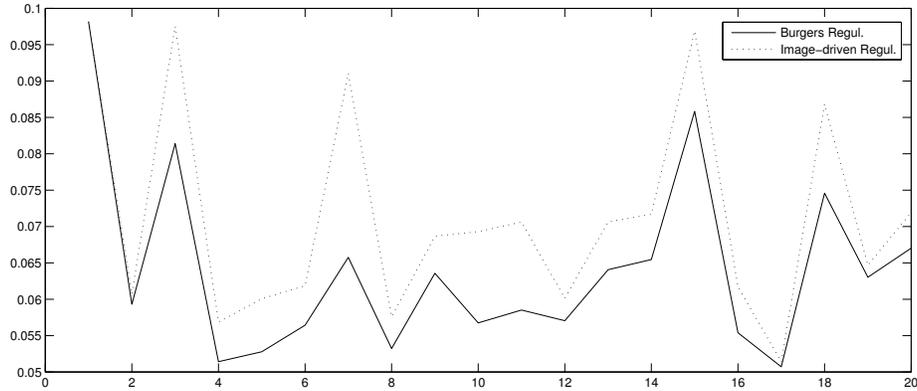15. R.J. LeVeque. *Numerical Methods for Conservation Laws*. Birkhäuser Verlag, 1990.

Fig. 8. Comparison of the L^2 motion error between the purely spatial image-driven variational approach (3) and the approach (1) together with (5). The latter combination allows for temporal regularization with a recursive computational architecture, leading to consistently lower estimation errors if actual and expected motions agree (parameters: $\alpha = 0.01$, $\beta = 0.001$).

16. G. Mather, F. Verstraten, and S. Anstis, editors. *The Motion Aftereffect.* MIT Press, 1998.

17. H.H. Nagel. On the estimation of optical flow: Relations between different approaches and some new results. *Artif. Intell.*, 33:299–324, 1987.

18. C. Schnörr. Determining optical flow for irregular domains by minimizing quadratic functionals of a certain class. *Int. J. Computer Vision*, 6(1):25–38, 1991.

19. J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in pde–based computation of image motion. *Int. J. Computer Vision*, 45(3):245–264, 2001.

20. J. Weickert and C. Schnörr. Variational optic flow computation with a spatio-temporal smoothness constraint. *J. Math. Imaging and Vision*, 14(3):245–255, 2001.

21. G.B. Whitham. *Linear and Nonlinear Waves.* Pure & Applied Mathematics, 1974.