IMAGE MOTION ESTIMATION USING OPTIMAL FLOW CONTROL

Annette Stahl * and Ole Morten Aamo
Department of Engineering Cybernetics, Norwegian University of Science and Technology (NTNU), Norway
anstahl@itk.ntnu.no, aamo@ntnu.no

Keywords: Motion estimation, Optimal control, Physical prior, Optimisation.

Abstract: In this paper we present an optimal control approach for image motion estimation in an explorative and novel way. The variational formulation incorporates physical prior knowledge by giving preference to motion fields that satisfy appropriate equations of motion. Although the framework presented is flexible, we employ the Burgers equation from fluid mechanics as physical prior knowledge in this study. Our control based formulation evaluates entire spatio-temporal image sequences of moving objects. In order to explore the capability of the algorithm to obtain desired image motion estimations, we perform numerical experiments on synthetic and real image sequences. The comparison of our results with other well-known methods demonstrates the ability of the optical control formulation to determine image motion from video and image sequences, and indicates improved performance.

1 INTRODUCTION

In this work we are concerned with motion estimation of objects in image sequences. The understanding and reconstruction of dynamic motion in image scenes is one of the key problems in computer vision and robotics. We present an attempt to adopt control methods from the field of applied mathematics in a new form to image sequence processing and to provide preliminary evaluations of the capability of this approach.

We describe motion as the displacement vector field of pixels between consecutive frames of an image sequence. In the literature this is known as optical flow (Jain et al., 1995). In computer vision local and global approaches are used to compute the optical flow field of image sequences. Local approaches are designed to compute the optical flow at a certain pixel position by using only the image information in the local neighbourhood of this specific pixel (Lucas and Kanade, 1981). Variational optical flow methods represent global optimisation problems which can be used to recover the flow field from an image sequence as a global minimiser of an appropriate energy functional. Usually, these energy functionals consist of two terms: a data term that imposes the result to be consistent with the measurement (here the brightness constancy assumption) and a regularisation term which imposes additional constraints like global or piecewise smoothness to the optical flow field.

One of the first variational methods for motion analysis was introduced by (Horn and Schunck, 1981) and incorporates a homogeneous regularisation term, where the optical flow is enforced to vary smoothly in space. This leads to an undesired blurring across motion discontinuities. Therefore, regularisation terms were introduced to regularise the flow in an image-driven (Schnörr, 1991; Alvarez et al., 1999) or flow-driven (Deriche et al., 1995) way, where the flow is prevented from smoothing across object or motion boundaries, respectively. A systematic classification of these approaches can be found in (Weickert and Schnörr, 2001a).

Most of the variational approaches incorporate a purely spatial regularisation of the flow. However, some efforts have been made to incorporate temporal smoothness (Nagel, 1990). The work of (Weickert and Schnörr, 2001b) investigates an extension of spatial flow-driven regularisation terms to spatio-temporal flow-driven regularisers. Time is considered as a third dimension analogue to the two spatial dimensions. These approaches improve both the robustness and the accuracy of the motion estimation but the
flow computation involves the data of the full image sequence at once.

Note that all these approaches do not incorporate physical prior knowledge about the motion itself. In contrast our approach incorporates a space-time regularisation using physical prior knowledge in a control framework that draws on the literature on the control of distributed parameter systems in connection with fluid dynamics (Gunzburger, 2002).

The ideas of two existing control approaches that are related to motion computation of image sequences are presented by (Ruhnau and Schnörr, 2007) and (Borzi et al., 2002). Ruhnau and Schnörr presented an optical flow estimation approach for particle image velocimetry that is based on a control formulation subject to physical constraints (Stokes equation). Their aim is to estimate the velocities of particles in image sequences of fluids rather than to estimate motion in every day image scenes.

The basic idea of (Borzi et al., 2002) is to estimate both an optical flow field \( u \) and a rectified image function \( I \) satisfying the brightness constancy assumption. Note that in their approach \( Y_k \) (and not \( I \)) denotes the sampled images of the image sequence. The most significant difference to our optical flow approach is that they do not only estimate the optical flow \( u \), but also \( I_k \), which is an approximation of the captured grey value distributions \( Y_k \), where \( k \) specifies the frame number within the image sequence. As part of the first-order necessary optimality conditions of the Lagrangian functional their optimal control formulation does not require a differentiation of the image data.

In contrast to that approach, we interpret the grey values of a scene as a "fictive fluid" - assuming that its motion can be described by an appropriate physical model, in this work realised with the Burgers equation of fluid mechanics. We adopt the well established variational optical flow approach of (Horn and Schunck, 1981) and add a distributed control exploiting the Burgers equation resulting in a constrained minimisation problem. The obtained objective functional has to be minimised with respect to the optical flow and control variables subject to the model equation over the entire flow domain in space and time. Our approach estimates not only the optical flow data from an image sequence, but it also estimates a force driven by the Burgers equation. The force field indicates the violation of the equation and can indicate accelerated motions like starting or stopping events or the change of the motion direction. Therefore one can exploit this feature as an indicator of unexpected motion events, taking place in the image sequence.

The initially constrained optimisation problem is reformulated - exploiting Lagrange multipliers - into an unconstrained problem allowing to obtain the associated first-order optimality system. This results in a forward-backward system with appropriate initial and boundary conditions. To solve the optimality system we uncouple the forward and backward computation as described in (Gunzburger, 2002) leading to an iterative solution scheme.

2 APPROACH

Before we start to describe the approach in more detail we first exemplify the notation and components of our control formulation.

We define a grey value of a certain pixel within an image sequence by a real valued one-time continuously differentiable \( C^1 \) image function \( I(x,t) \), where \( x = (x_1,x_2)^T \) denotes the location within some rectangular image domain \( \Omega \) and \( t \in [0,T] \) labels the corresponding frame at time \( t \). In particular, the function \( I(x_1,x_2,t) \) denotes the intensity of a pixel at position \( (x_1,x_2)^T \) in the image frame at time \( t \). The optical flow field is denoted by a two-dimensional vector field \( u = (u_1(x,t),u_2(x,t))^T \), which describes the intensity changes between images.

We formulate our motion estimation problem within a variational framework. We minimise an energy functional \( E \), which consists of a data and a regularisation term:

**Data Term.** We make use of the following data term

\[
\int_{\Omega} (\partial_t I + u \cdot \nabla I)^2 \, dx, \quad (1)
\]

which comprises the optical flow constraint (Horn and Schunck, 1981) and provides the link between the given image data, the observed intensity \( I \) and the desired velocity field \( u \). Note that the optical flow constraint equation represents the requirement that the intensity of an object point stays constant along its motion trajectory. Problem (1) is ill-posed as any vector field \( u \) satisfying \( u \cdot \nabla I = -\partial_t I \) is a minimiser. Therefore a regularisation term is added to introduce additional constraints for the flow field \( u \) to obtain an unique solution.

**Regularisation Term.** We incorporate the regularisation term from (Horn and Schunck, 1981)

\[
\int_{\Omega} \alpha(|\nabla u_1|^2 + |\nabla u_2|^2) \, dx, \quad 0 < \alpha \in \mathbb{R}, \quad (2)
\]

to enforce spatial smoothness of the optical flow field, preferring neighbouring optical flow vectors to...
be similar. The regularisation parameter $\alpha$ adjusts the relative importance of the smoothness term to the data term. With an increasing value of $\alpha$ the vector field is forced to become smoother. We are aware that regulariser like the L1-regulariser used for example in (Wedel et al., 2009) allows for sharper discontinuities in the flow field. Our decision to use the L2-regulariser in the motion estimation was mainly driven by the idea to keep the approach clear and numerically simple. However, the replacement of the quadratic homogeneous smoothness term could improve the accurateness of the computed motion boundaries.

**Physical Prior.** Considering a constant moving object one can determine that structures are transported by a velocity field and along with it the velocity field is transported by itself. A physical model equation, which describes this behaviour is the Burgers equation and allows to model the movement of rigid objects.

The inviscid Burgers equation

$$\frac{Du}{Dt} = \partial_t u + (u \cdot \nabla) u = 0, \quad u(x, 0) = u_0$$

(3)

has been studied and successfully applied for many decades in aero- and fluid dynamics (Burgers, 1948; Hirsch, 2000) as a simplified model for turbulence, boundary layer behaviour, shock wave formation and mass transport. It contains the convection term from the fundamental equations of fluid mechanics, the Navier-Stokes equations.

As a physical interpretation, $u$ in (3) 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 $\frac{Du}{Dt}$ yields the acceleration of moving particles. The nonlinear term $(u \cdot \nabla) u$ is known as the inertia term of the transport process described by (3). See Figure 1 for an illustration of the transport. We found that our approach even with the constant velocity assumptions of our physical prior predicts the non-uniform motion pattern quite well as shown in our numerical results (cf. Sec. 4.2).

### 2.1 Optimal Control Formulation

In the following sections we explain our optimal control approach. Foundations exploiting fluid dynamical methods can be found in the book of Gunzburger (Gunzburger, 2002).

We obtain our spatial-temporal control approach as follows: Additionally to the smoothness term we introduce a control $f$, that is distributed in space and time, which means that it acts over the entire optical flow domain $\Omega \times [0, T]$. The magnitude of the control is bounded due to penalisation within the objective functional. The resulting optimisation problem is to minimise

$$E(u, f) = \frac{1}{2} \int_{\Omega \times [0, T]} \left\{ (\partial_t I + u \cdot \nabla I)^2 \right\} dx dt,$$

subject to the equations of motion

$$\begin{cases} \partial_t u + (u \cdot \nabla) u = f & \text{in } (0, T) \times \Omega, \\ \partial_x u = 0 & \text{on } (0, T) \times \Gamma. \end{cases}$$

(5)

We intend to find an optimal state $u = (u_1, u_2) ^T$ and an optimal control $f = (f_1, f_2) ^T$, such that the functional $E(u, f)$ is minimised and $u$ and $f$ satisfy the Burgers equation (5).

The objective of this formulation is to determine a body force $f$ (the control !) that leads to a velocity field $u$ which fits to the apparent motion in the image sequence, and at the same time satisfies physical prior knowledge in terms of the given equations of motion.

### 2.2 Optimality System

In order to obtain the velocity field $u$ and the control $f$ we recast the constrained optimisation problem (4) - (5) into an unconstrained optimisation problem. Introducing the Lagrange multiplier or adjoint variable $w = (w_1(x, t), w_2(x, t)) ^T$ yields the following Lagrangian functional

$$L(u, f, w) = E(u, f)$$

$$- \int_{\Omega \times [0, T]} w^T (\partial_t u + (u \cdot \nabla) u - f) dx dt.$$

(6)

To solve this functional we have to derive the first-order necessary conditions. This results in the following optimality system (7)-(9) from which the optimal
state $u$, adjoint $w$, and the optimal control $f$ can be determined such that $L(u, f, w)$ is rendered stationary.

\[
\begin{align*}
\partial_t u + (u \cdot \nabla) u &= f \quad \text{in } \Omega \times [0, T], \\
\partial_t u &= 0 \quad \text{on } \Gamma \times [0, T], \\
\begin{cases}
\partial_t w - (u \cdot \nabla) w - w\nabla \cdot u + (\nabla U)^T w \\
= \nabla I(\partial_t u + u \cdot \nabla I) - \alpha \Delta u
\end{cases} \quad \text{in } \Omega \times [0, T], \\
w|_{t=0} &= 0 \quad \text{in } \Omega, \\
\beta f + w &= 0 \quad \text{in } \Omega \times [0, T], \\
f|_{t=0} &= 0 \quad \text{on } \Gamma \times [0, T], \\
f|_{t=T} &= 0 \quad \text{in } \Omega,
\end{align*}
\]

where $(\nabla U)^T$ the transposed Jacobian matrix.

The state equation (7) is obtained by derivation of the Lagrangian functional (6) in the direction of the Lagrange multiplier, and out to be identical to the Burgers equation (5) itself. The adjoint equation (8) specifies the first-order necessary conditions with respect to the state variables $u$. The optimality condition (9) is the necessary condition that the gradient of the objective function – with respect to the control $f$ – vanishes at the optimum. It also includes the initial and terminal conditions.

The optimality system (7)-(9) is a coupled system which turns out to be - due to the large number of unknowns - prohibitively expensive to solve directly, but can be solved iteratively as described in the next section.

### 2.3 Algorithm

We solve the optimality system (7)-(9) using an iterative gradient descent method (with step length adaptation) which decouples the state and adjoint computation. It consists of the iterative solution of the state and adjoint equation in such a way that the state equation is computed forward in time with appropriate initial condition $u_0$ and the adjoint equation is computed backward in time with terminal condition $w|_{t=T} = 0$. The optimality condition is used to update the control $f$ with the adjoint variable $w$. The control $f$ is then used to compute the actual state $u$. Additionally, the step length is adjusted ensuring that the actual energy of the objective functional (4) decreases. Note that we choose the start value for $f$ to be zero in the very first iteration.

In our pseudo code description of Algorithm 1, variable $s$ denotes the step-size that is adapted by the algorithm and $\varepsilon$ the threshold which is used to decide if the relative difference of the energy is small enough to be seen as converged.

In the initial step of the algorithm the flow fields $u$ for all consecutive image frames and the terminal condition of the adjoint variable for the last frame ($w|_{t=T}$) are set to zero. The first step of the iteration loop solves the adjoint equation (8) for $w$ backwards in time using the terminal condition on $w$ and the flow field $u$. Then, the optimality condition (9) is used to update the control field for all frames, allowing the state equation (8) to be solved for $u$ forward in time using the new control field. The iteration loop continues until the decline in $E$ is negligible.

**Algorithm 1: Gradient algorithm with automatic step-length selection.**

1: set $u = 0$, $\varepsilon = 10^{-8}$, and $s := s_0$ (initial step) 
2: repeat 
3: solve the adjoint equation (8) for $w$ 
4: update $f$: $f_m = f_{m-1} - s(\beta f_{m-1} + w)$ 
5: solve the state equation (7) for $u$ 
6: if $E(u, f_m) \geq E(u, f_{m-1})$ then 
7: $s := 0.5s$ 
8: goto 4 
9: else 
10: $s := 1.5s$ 
11: end if 
12: until $|E(u, f_m) - E(u, f_{m-1})|/|E(u, f_m)| < \varepsilon$

### 3 NUMERICAL SOLUTION

In this part, we summarised the numerical discretisation methods employed in solving the optimality system (7)-(9). For more details, we refer to (Colella and Puckett, 1998).

**Discretisation of the State Equation.** Within the numerical implementation of the nonlinear state system equation (7) we have to cope with overlapped undershoots, with shock formations, with the compliance of conditions (entropy-, monotony-, CFL-condition, etc.) and different discretisation schemes. We use the second-order conservative Godunov scheme for our implementation. The fluxes are numerically computed by solving the equations at pixel edges. The correct behaviour at discontinuities is obtained by using solutions of the appropriate Riemann problem.

**Discretisation of the Adjoint Equation.** The numerical implementation of the time-dependent adjoint system (8) in the domain $\Omega$ is done by using a second-order predictor-corrector finite difference
scheme. The basic idea behind this is that all methods with an accuracy larger than the order one will produce spurious oscillations in the vicinity of large gradients, while being second-order accurate in regions where the solution is smooth. To prevent such oscillations the slopes of Fromm’s method are replaced by the slopes of the Van Leer scheme. The Van Leer scheme detects discontinuities and modifies its behaviour in such locations accordingly. The implication of this is that this method retains the high-order accuracy of Fromm’s scheme in smooth regions, but near discontinuities the discretised evolution equation drops to first-order accuracy.

4 EXPERIMENTS

In this section we first illustrate the control performance of our optical flow approach on a real-world 2D image sequence. Secondly, we evaluate the following motions which violate the incorporated motion assumption: rotation, translation in combination with scaling. Finally, we present the results for noisy image data showing the influence of the temporal regularisation in the control approach and provide a comparison with error measures obtained by the approach from (Horn and Schunck, 1981) and the dynamic optical flow approach from (Stahl et al., 2006).

4.1 Control - Force

We illustrate the control behaviour of our approach for a real-world 2D image sequence with an unexpected motion. The image sequence consists of 10 image frames and shows a moving hand which starts to move and then stops again. Figure 2 depicts the starting (left column) and stopping (right column) event of the sequence. The first row shows the velocity estimates $u$, and the second row shows the force fields. The force field $f$ nicely indicates the deviation of the expected motion from the observed motion. This is evident in the second row of Figure 4, where the force field acts in the direction of the moving hand as the hand accelerates into motion (left picture), while it turns in the opposite direction as the hand stops (right picture).

4.2 Non-uniform Motion

In this section we provide an evaluation of our approach on the basis of two well known synthetic image sequences for which the ground truth motion data is available. To allow for a quantitative comparison we provide the results we obtain for the Horn and Schunck as well. The image sequences we use show global motion patterns such as rotation, translation and divergence.

In particular we evaluate our approach on the grey value versions of the following two image sequences: the “rotating sphere” sequence (McCane et al., 2001) and the ”Yosemite” sequence (available at ftp://ftp.csd.uwo.ca/pub/vision).

The “rotating sphere” sequence contains a curling vector field and is shown in Figure 3. This sequence consists of 45 frames, where a sphere rotates in front of a stationary background.
The computed vector fields obtained by the Horn and Schunck approach and our approach (4)-(5) for the “rotating sphere” sequence is shown in Figure 4.

Figure 4: The synthetic “rotating sphere” sequence. Computational Results for the Horn and Schunck approach and the control based approach. Left: Result Horn and Schunck (RMSE = 0.395). Right: Result control based approach (RMSE = 0.192).

The motion estimation results for the “Yosemite” sequence are shown in section 4.3.

The results show that even for sequences which violate the constant velocity assumption of the model equation we obtain good results. However, due to the flexibility of our variational approach it should be possible to model such motion patterns by incorporation of a suitable model equation.

4.3 Temporal Regularisation

To investigate the impact of the temporal regularisation to the robustness of our approach under noise, we choose the “Yosemite” sequence with different Gaussian noise levels $\sigma = 0, 10, 20$ and 40 (cf. Fig. 5). The sequence exhibits divergent and translational motion combined with illumination changes. To investigate the performance of our approach we compare the root mean square error (RMSE)

$$RMSE(u_o, u_e) = \frac{1}{|\Omega|} \int_{\Omega} \sqrt{(u_o - u_e)^2} \, dx$$

and the average angular error (AAE)

$$AAE(u_o, u_e) = \frac{1}{|\Omega|} \int_{\Omega} \arccos \left( \frac{u_o \cdot u_e}{|u_o||u_e|} \right) \, dx,$$

where $\cdot \cdot$ denotes the Euclidean norm, $u_o = (u_{o1}, u_{o2}, 1)^T$ the original optical flow vectors, and $u_e = (u_{e1}, u_{e2}, 1)^T$ the estimated optical flow vectors (compare (Barron et al., 1994)). Note that the time dimension is set to 1 corresponding to the distance of one frame.

This measure is currently used as a kind of standard to provide accuracy measures for optical flow results.

We compare the errors of the optical flow computation obtained for three different approaches with optimised parameters. In particular these are the homogeneous spatial regularised approach from (Horn and Schunck, 1981), the spatio-temporal dynamic image motion approach from (Stahl et al., 2006), and our control based image motion approach (4) - (5). The control approach results in a improved vector field, which is based on the forward-backward computation, which incorporates additional knowledge of the future frames leading to an improved temporal regularisation. The result for a single frame in the highly noisy Yosemite sequence is shown in Figure 6. The results for the computed errors (RMSE and AAE) for all three approaches with increasing noise level are shown in Table 1. The purely spatial regularised approach from Horn and Schunck and the absence of physical prior knowledge leads to the higher error values with increasing noise levels. In contrast to the spatio-temporal dynamic image motion approach (Stahl et al., 2006) a higher noise level requires the selection of a smaller $\beta$ regularisation parameters for the control part of the objective functional. The consistently lower error indicates an improved global motion prediction in our control approach (4)-(5) exerting a better temporal regularisation. Our explanation for this observation is that the control approach incor-
Table 1: Performance of our control approach (C) in comparison with the Horn and Schunck approach (HS) and the dynamic image motion approach (Dy) in presence of noise: We added random Gaussian noise with zero mean and standard deviation $\sigma = 0, 10, 20, \text{and} 40$ to the Yosemite image sequence.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>app.</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>RMSE</th>
<th>AAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>HS</td>
<td>0.005</td>
<td>-</td>
<td>0.177</td>
<td>3.04°</td>
</tr>
<tr>
<td></td>
<td>Dy</td>
<td>0.006</td>
<td>0.00002</td>
<td>0.178</td>
<td>3.09°</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.007</td>
<td>0.0005</td>
<td>0.169</td>
<td>2.88°</td>
</tr>
<tr>
<td>10</td>
<td>HS</td>
<td>0.008</td>
<td>-</td>
<td>0.283</td>
<td>5.74°</td>
</tr>
<tr>
<td></td>
<td>Dy</td>
<td>0.01</td>
<td>0.0003</td>
<td>0.275</td>
<td>5.68°</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.009</td>
<td>0.0001</td>
<td>0.243</td>
<td>4.92°</td>
</tr>
<tr>
<td>20</td>
<td>HS</td>
<td>0.02</td>
<td>-</td>
<td>0.429</td>
<td>8.61°</td>
</tr>
<tr>
<td></td>
<td>Dy</td>
<td>0.025</td>
<td>0.001</td>
<td>0.395</td>
<td>7.54°</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.02</td>
<td>0.00001</td>
<td>0.350</td>
<td>6.67°</td>
</tr>
<tr>
<td>40</td>
<td>HS</td>
<td>0.05</td>
<td>-</td>
<td>0.640</td>
<td>13.27°</td>
</tr>
<tr>
<td></td>
<td>Dy</td>
<td>0.05</td>
<td>0.005</td>
<td>0.523</td>
<td>9.89°</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.05</td>
<td>0.000003</td>
<td>0.497</td>
<td>9.16°</td>
</tr>
</tbody>
</table>

5 CONCLUSIONS

We have presented an optimal control approach to image motion estimation including physical prior knowledge in a novel and exploratory way. It leads to an unconstrained optimisation problem, where the optimality system - from which the optimal state and the optimal control are determined - can be solved using an iterative gradient descent method. The forward-backward structure of the model allows for a robust estimation of the coherent flows by including prior knowledge that enforce spatio-temporal smoothness of the minimising vector field.

In the case that the image measurements indicate changes of the current velocity distribution, fictive control forces modify the system state accordingly. The presence of such forces may serve as an indicator notifying a higher-level processing stage about unexpected motion events in video sequences.

The comparison of our results with the approach from (Horn and Schunck, 1981) and the approach from (Stahl et al., 2006) demonstrates the ability of the control formulation to determine image motion from video sequences, and shows improved performance, especially for highly noisy image data. Our further work will include the modification of the Burgers equation to achieve better motion boundaries in the rarefaction area and the reformulation of the approach to a receding horizon formulation.

ACKNOWLEDGEMENTS

We would like to thank Dr. Christian Schellewald, Dr. Paul Ruhnau, Prof. Christoph Schnörr, and Prof. Øyvind Stavdahl for some inspiring discussions and comments.

REFERENCES


