

# Modeling Temporal Coherence for Optical Flow

Sebastian Volz, Andrés Bruhn, Levi Valgaerts  
Vision and Image Processing Group  
Saarland University, Germany  
`{volz,bruhn,valgaerts}@mmci.uni-saarland.de`

Henning Zimmer  
Mathematical Image Analysis Group  
Saarland University, Germany  
`zimmer@mia.uni-saarland.de`

## Abstract

*Despite the fact that temporal coherence is undeniably one of the key aspects when processing video data, this concept has hardly been exploited in recent optical flow methods. In this paper, we will present a novel parametrization for multi-frame optical flow computation that naturally enables us to embed the assumption of a temporally coherent spatial flow structure, as well as the assumption that the optical flow is smooth along motion trajectories. While the first assumption is realized by expanding spatial regularization over multiple frames, the second assumption is imposed by two novel first- and second-order trajectory smoothness terms. With respect to the latter, we investigate an adaptive decision scheme that makes a local (per pixel) or global (per sequence) selection of the most appropriate model possible. Experiments show the clear superiority of our approach when compared to existing strategies for imposing temporal coherence. Moreover, we demonstrate the state-of-the-art performance of our method by achieving Top 3 results at the widely used Middlebury benchmark.*

## 1. Introduction

Temporal coherence plays an important role in many computer vision tasks based on video data. By assuming the results at the current time instant to be similar to the ones observed at previous time instants, the estimation process can be regularized and the results typically become more stable. While this concept forms the basis of many tracking algorithms, it is surprising that it has hardly entered the field of motion estimation so far. In fact, the currently best performing optical flow methods are based on energy formulations and should thus allow for a transparent modeling of temporal coherence. However, most of them focus either on incorporating more advanced data constancy assumptions to improve robustness under noise and varying illumination [5, 17, 23, 25] or spend significant efforts on the design of discontinuity-preserving smoothness constraints to ensure *spatial* coherence [17, 21, 23, 26]. If at all, temporal coherence is only enforced afterwards by computing trajectories from previously estimated flow fields [16, 19]. The absence of temporal coherence in recent models also becomes evident when taking a look at the Middlebury benchmark [2]: Only two out of 48 methods make use of more than two frames and none of them incorporates temporal information in the underlying model. Nevertheless, there have been some attempts in the literature to integrate such information. As we will see from the following survey, these techniques can be classified into three groups that all suffer from specific shortcomings:

ence is only enforced afterwards by computing trajectories from previously estimated flow fields [16, 19]. The absence of temporal coherence in recent models also becomes evident when taking a look at the Middlebury benchmark [2]: Only two out of 48 methods make use of more than two frames and none of them incorporates temporal information in the underlying model. Nevertheless, there have been some attempts in the literature to integrate such information. As we will see from the following survey, these techniques can be classified into three groups that all suffer from specific shortcomings:

**I. Temporal Regularization.** Early ideas go back to Murray and Buxton [13] who introduced a spatio-temporal smoothness term that assumes the resulting flow fields to be *smooth along the temporal axis*. However, in practice, such an assumption hardly makes sense, since moving objects naturally change their location over time. Nagel [14], and later Zimmer *et al.* [26], alleviated this problem by using directional smoothness assumptions in the space-time volume. Similarly motivated, Weickert and Schnörr [22] suggested a robust spatio-temporal formulation to tackle the aforementioned problem of temporal inconsistency. The latter idea was then extended to motion segmentation with spatio-temporal level sets by Brox *et al.* [6]. However, all these methods suffer from a common problem: In the presence of larger displacements, the temporal derivatives in the smoothness term do not allow to capture the actual trajectories of objects. Similar problems also arise in the methods of Chin *et al.* [10] and Elad and Feuer [11] that are based on Kalman filtering with a temporally constant motion model.

**II. Trajectory Stabilization.** A completely different approach was proposed by Black and Anandan [3]. They suggested to additionally compute flow fields between *previous frames* and register them to the current time instant. These flows are then used within a robust *similarity prior* that guides the estimation of the current flow field. While this strategy is suitable for larger displacements, flow information is only propagated forward in time. Moreover, there is no feedback between the different flow fields during the computation, since they are not estimated simultaneously.

**III. Trajectorial Regularization.** Most suitable from a modeling viewpoint are smoothness constraints that assume the flow fields to be *smooth along the trajectories* of moving objects. However, there are only a few works in the literature that address this kind of regularization. Werlberger *et al.* [24] consider three frames and enforce trajectorial constancy as a *hard constraint* by parametrizing the two flow fields by a single function. Since it is very likely that objects change orientation and speed over time, it is not surprising that the authors report a deterioration of the results in most cases. More appropriately, Salgado and Sanchez [15] propose a trajectorial regularizer that penalizes first-order variations along motion trajectories and thus acts as a *soft constraint*. While their approach allows to use arbitrarily many frames and offers a higher flexibility than the constant model in [24], it still suffers from two main drawbacks: On the one hand, the optimization is rather difficult, since flow fields at different time instants refer to different coordinate frames and thus have to be repeatedly registered onto each other. On the other hand, the results become significantly worse in the presence of more complex motion patterns, since the trajectorial regularization oversmoothes the results in those cases. Similar in spirit is the work of Chaudhury and Mehrotra [9] that assumes the curvature of motion trajectories to be small. While this assumption is even more general than the first-order model from [15], the underlying differential model limits the reliable estimation to small displacements. Moreover, the optimization is difficult to handle and thus realized via a generic algorithm (simulated annealing). Also in this case, more complex motion patterns result in a significant reduction of accuracy.

**Contribution.** We tackle all of the aforementioned problems by four contributions: (i) First, we propose a *novel parametrization along motion trajectories* for computing the optical flow from multiple input images. Since all flow fields then naturally become registered onto one reference frame, the cumbersome registration of flows during the optimization can be avoided. (ii) Based on this parametrization we suggest a joint spatial regularization of all flows at different time instants. This allows us to model the assumption of a *temporally coherent spatial flow structure*. (iii) We extend our model by a purely trajectorial regularization of first and second order. This allows us to obtain flow fields that are *pointwise temporally coherent*. (iv) Finally, we suggest an estimation strategy that allows us to *adapt the degree of trajectorial regularization* depending on the whole image sequence (global) or the pixel location (local). This is important in the presence of motion patterns with varying order of trajectorial smoothness or where such smoothness does not hold. A comparison to methods with spatial and temporal regularization as well as trajectorial stabilization shows the clear superiority of our approach. This performance is confirmed by Top 3 results in the Middlebury benchmark.

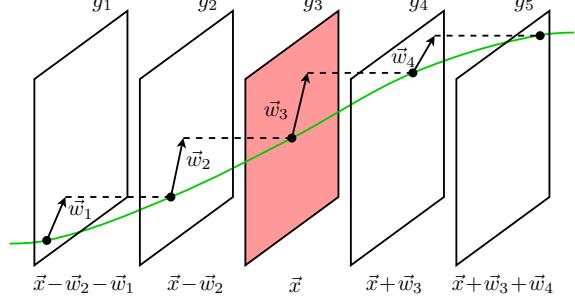


Figure 1. Temporal scenario with flow trajectory over five frames.

**Organization.** In Section 2 we derive our model for multi-frame optical flow computation with trajectorial parametrization. This model is then extended by adaptive first- and second-order trajectorial regularization in Section 3. Section 4 is dedicated to minimization issues, while our results and a comparison to the literature are presented in Section 5. We conclude with a summary in Section 6.

## 2. A Model with Trajectorial Parametrization

For our model, we consider five consecutive frames of a given image sequence. We assume that these frames have previously been convolved with a Gaussian  $K_\sigma$  of standard deviation  $\sigma$  to eliminate noise and undesired high frequencies. As illustrated in Fig. 1, we denote the five images by  $g_1(\vec{x}), \dots, g_5(\vec{x})$ , where  $\vec{x} := (x, y)^\top$  is the location within a rectangular image domain  $\Omega \subset \mathbb{R}^2$ . Consequently, we can define four optical flow fields  $\vec{w}_1, \vec{w}_2, \vec{w}_3$  and  $\vec{w}_4$ , where the  $i$ -th flow  $\vec{w}_i(\vec{x}) := (u_i(\vec{x}), v_i(\vec{x}))^\top$  describes the displacement between the images  $g_i$  and  $g_{i+1}$ . Our goal in this scenario is now to estimate the flow field  $\vec{w}_3$  between the central frame and its successor. In particular, we aim at improving its accuracy by simultaneously estimating the three other flows as well. To this end, we propose a variational approach that minimizes an energy of the general form

$$\mathcal{E} = \int_{\Omega} (\mathcal{E}_d + \mathcal{E}_s + \mathcal{E}_t) \, d\vec{x}. \quad (1)$$

Here,  $\mathcal{E}_d$  is a data term that models constancy assumptions on image features and  $\mathcal{E}_s$  and  $\mathcal{E}_t$  are smoothness terms that enforce the flows to be smooth in spatial and trajectorial directions, respectively.

### 2.1. Reference Frame and Parametrization

Using the standard parametrization for spatio-temporal methods, where each optical flow field  $\vec{w}_i$  is referenced with respect to the frame  $g_i$  (e.g. [5, 14, 26]) makes a propagation of information along motion trajectories difficult. In fact, it requires to repeatedly register flow fields from different time instants onto each other during the minimization [15].

Our solution to this problem is to designate a *single reference frame* in the sequence and to parametrize all flow fields w.r.t. this frame; see Fig. 1. Since we are interested in computing  $\vec{w}_3$ , the central image  $g_3$  will be our reference frame. This is indicated by the red color in the sketch. Note that due to the chosen parametrization, the flow  $\vec{w}_4$  for example does not directly describe the displacement from  $g_4$  to  $g_5$ , but gives the flow increment that has to be added to  $\vec{w}_3$  in order to obtain the trajectory from  $g_3$  to  $g_5$ . Using the described parametrization thus offers the advantage that all flow fields refer to the same coordinate system. In particular, we can model temporal coherence of the flows along their trajectories by simply assuming that the flow vectors in the same reference location should be similar. Furthermore, we can also expect that the discontinuities of all flow fields occur at the same reference locations. In the modeling of the spatial and trajectory smoothness term we will exploit both findings.

Before turning to the actual design of the individual terms in our energy (1), let us remark that our model is in general applicable to arbitrarily many frames. In this paper, however, we focus on the five frame case, since it allows us to propose the concept of higher-order trajectory regularization while keeping the model tractable at the same time.

## 2.2. Data Constraints

We first derive the four constraints that model the relations between all image pairs  $g_i$  and  $g_{i+1}$ . For simplicity, let us assume for the moment that the brightness of corresponding image points remains constant between all frames. Following the notation in Fig. 1 we then obtain the four expressions:

$$\mathcal{E}_{d12} = \theta \cdot \Psi_d (|g_2(\vec{x} - \vec{w}_2) - g_1(\vec{x} - \vec{w}_2 - \vec{w}_1)|^2), \quad (2)$$

$$\mathcal{E}_{d23} = \Psi_d (|g_3(\vec{x}) - g_2(\vec{x} - \vec{w}_2)|^2), \quad (3)$$

$$\mathcal{E}_{d34} = \Psi_d (|g_4(\vec{x} + \vec{w}_3) - g_3(\vec{x})|^2), \quad (4)$$

$$\mathcal{E}_{d45} = \theta \cdot \Psi_d (|g_5(\vec{x} + \vec{w}_3 + \vec{w}_4) - g_4(\vec{x} + \vec{w}_3)|^2). \quad (5)$$

We use the regularized  $L_1$  norm

$$\Psi_d(s^2) = \sqrt{s^2 + \epsilon_d^2}, \quad (6)$$

with  $\epsilon_d = 0.001$  as sub-quadratic penalizer function in order to render our method robust against outliers caused by noise or occlusions [4]. As outliers may occur independently in the four data constraints, we apply the penalizer separately to each brightness difference. We also introduce a weight  $\theta = 0.5$  in order to downweight the influence of constraints that have a larger temporal distance to the reference frame. Such constraints typically contain larger motion (a sum of flows) and should thus be considered less reliable for the estimation.

Additionally, we impose the gradient constancy assumption to cope with illumination changes [5, 7] and extend the data constraints to RGB color images. Incorporating these modifications, the data term  $\mathcal{E}_{d34}$  for instance becomes

$$\begin{aligned} \mathcal{E}_{d34} = & \Psi_d \left( \sum_{c=1}^3 (|g_4^c(\vec{x} + \vec{w}_3) - g_3^c(\vec{x})|^2) \right) \\ & + \gamma \Psi_d \left( \sum_{c=1}^3 (|\nabla g_4^c(\vec{x} + \vec{w}_3) - \nabla g_3^c(\vec{x})|^2) \right), \end{aligned} \quad (7)$$

where  $\gamma \geq 0$  is a weighting factor, the symbol  $\nabla = (\partial_x, \partial_y)^\top$  denotes the spatial gradient operator, and  $g^c$  (for  $c=1, 2, 3$ ) represents the three RGB color channels. The remaining data constraints  $\mathcal{E}_{d12}$ ,  $\mathcal{E}_{d23}$  and  $\mathcal{E}_{d45}$  are extended accordingly. By summing up all data constraints, we obtain

$$\mathcal{E}_d = \mathcal{E}_{d12} + \mathcal{E}_{d23} + \mathcal{E}_{d34} + \mathcal{E}_{d45}, \quad (8)$$

which forms the overall data term of our final model.

## 2.3. Registered Complementary Regularization

Next, we turn to the design of the spatial smoothness term  $\mathcal{E}_s$ . It penalizes fluctuations in the flow fields and allows to regularize the flow estimation in locations where the data constraints are insufficient, e.g. in flat image regions or in the presence of outliers caused by noise or occlusions.

Our smoothness term is based on the *complementary regularizer* proposed in [26]. This anisotropic smoothness term relaxes the smoothness assumption in direction of the data constraints to avoid undesirable interference between the data and the smoothness term. To obtain the data constraint direction it is proposed to consider the eigenvectors of a regularization tensor that is computed from gradient information of the first image. Since in our method all flow fields are registered w.r.t. a single reference image, it also makes sense to compute a *single regularization tensor*  $R_\rho$ . This tensor is obtained from the reference image  $g_3$  as

$$\begin{aligned} R_\rho = \sum_{c=1}^3 K_\rho * & \left[ \nabla g_3^c \nabla g_3^{c\top} \right. \\ & \left. + \gamma \left( \nabla g_{3x}^c \nabla g_{3x}^{c\top} + \nabla g_{3y}^c \nabla g_{3y}^{c\top} \right) \right], \end{aligned} \quad (9)$$

where subscripts denote partial derivatives and  $\rho$  is a spatial integration scale. The orthonormal eigenvectors  $\vec{r}_1$  and  $\vec{r}_2$  of the regularization tensor  $R_\rho$  give the desired directional information as  $\vec{r}_1$  points in data constraint direction. This directional information is incorporated into the smoothness term by considering the projections of the flow gradients  $\nabla u$  and  $\nabla v$  onto  $\vec{r}_1$  and  $\vec{r}_2$ . In order to obtain a complementary regularization behavior, the projections in data constraint direction ( $\vec{r}_1$ -direction) should be penalized less severely than the projections onto  $\vec{r}_2$ .

## 2.4. Separate Spatial Regularization

Modeling such complementary regularizers for all four flow fields *separately* and combining them into two directional smoothness terms in  $\vec{r}_1$ - and in  $\vec{r}_2$ -direction then gives

$$\mathcal{E}_{s_1}^{\text{sep}} = \sum_{i=1}^4 \nu_i \Psi_{s_1} \left( (\vec{r}_1^\top \nabla u_i)^2 + (\vec{r}_1^\top \nabla v_i)^2 \right), \quad (10)$$

$$\mathcal{E}_{s_2}^{\text{sep}} = \sum_{i=1}^4 \nu_i \Psi_{s_2} \left( (\vec{r}_2^\top \nabla u_i)^2 + (\vec{r}_2^\top \nabla v_i)^2 \right). \quad (11)$$

Here, the weights  $\nu_i$  are introduced to ensure that the relative balance between the data and the smoothness constraints is the same for all flows. They are defined in accordance with the weight  $\theta$  in the data term: As  $\vec{w}_1$  and  $\vec{w}_4$  occur solely in a single data term weighted by  $\theta$ , we set  $\nu_1 = \nu_4 = \theta$ . In contrast, the flows  $\vec{w}_2$  and  $\vec{w}_3$  occur in two data terms that are weighted by 1 and  $\theta$ , respectively. Consequently, we chose  $\nu_2 = \nu_3 = 1 + \theta$  in this case.

## 2.5. Joint Spatial Regularization

If we recall that the parametrization of all flows w.r.t. the reference frame implies that flow discontinuities will appear simultaneously at the same location in all four flow fields, we can even go one step further: Instead of just summing up separate complementary regularizers for all flow fields, we can integrate all flows in two smoothness terms that encourage *joint spatial discontinuities* along motion trajectories:

$$\mathcal{E}_{s_1}^{\text{joint}} = \Psi_{s_1} \left( \sum_{i=1}^4 \nu_i \left( (\vec{r}_1^\top \nabla u_i)^2 + (\vec{r}_1^\top \nabla v_i)^2 \right) \right), \quad (12)$$

$$\mathcal{E}_{s_2}^{\text{joint}} = \Psi_{s_2} \left( \sum_{i=1}^4 \nu_i \left( (\vec{r}_2^\top \nabla u_i)^2 + (\vec{r}_2^\top \nabla v_i)^2 \right) \right). \quad (13)$$

Evidently, these smoothness terms are much more than pure spatial regularizers. Actually, they model the assumption of a *temporally coherent spatial flow structure*. A somewhat related assumption, but in the context of layered motion models, was formulated by Sun *et al.* [18]. There, the flow fields are decomposed into layered segments which are assumed to be coherent over two successive frames.

## 2.6. Choice of Spatial Penalizer Functions

Considering the choice of the penalizer functions  $\Psi$ , we realize the desired reduced smoothing behavior in  $\vec{r}_1$ -direction (the data constraint direction) by using the Perona-Malik penalizer (Lorentzian) [4] to define  $\Psi_{s_1}$  as

$$\Psi_{s_1}(s^2) = \lambda_1^2 \log \left( 1 + \frac{s^2}{\lambda_1^2} \right), \quad (14)$$

with a contrast parameter  $\lambda_1 = 0.1$ . In the orthogonal direction, we opt *against* a strong quadratic penalization as

originally proposed in [26]. Instead, we also use a sub-quadratic penalizer function, but one that still results in a stronger smoothing compared to the Perona-Malik function. Such a function is given by the Charbonnier penalizer

$$\Psi_{s_2}(s^2) = 2 \lambda_2^2 \sqrt{1 + \frac{s^2}{\lambda_2^2}}, \quad (15)$$

with a contrast parameter  $\lambda_2 = 0.1$ . Our experiments have shown that this choice can lead to qualitatively better results compared to the original quadratic penalizer: We observed sharper boundaries and less undesirable rounding at flow corners. Our final smoothness term is obtained by combining the two joint smoothness terms (12) and (13) as

$$\mathcal{E}_s = \alpha (\mathcal{E}_{s_1}^{\text{joint}} + \mathcal{E}_{s_2}^{\text{joint}}), \quad (16)$$

where  $\alpha > 0$  is a weight that balances the spatial smoothness assumptions against the other terms.

## 3. Adaptive Trajectorial Regularization

After having modeled the assumption of a temporally coherent spatial flow structure in the previous section, let us now discuss how we can additionally encourage *temporal smoothness along motion trajectories*.

### 3.1. Trajectorial Smoothness Constraints

To this end, we make once again use of the fact that all flow fields are registered w.r.t. one reference frame. This allows us to express the assumption of similar flows along the motion trajectory by the following regularizer:

$$\mathcal{E}_t^{\text{1st}} = \sum_{i=1}^3 \Psi_t \left( (u_{i+1} - u_i)^2 + (v_{i+1} - v_i)^2 \right). \quad (17)$$

From a numerical viewpoint, the two squared expressions in the argument of this space-continuous time-discrete term can be seen as a finite difference approximation of the first derivative of the flow components  $u$  and  $v$  in *trajectorial direction*. Similarly, if we use finite difference approximations of the second derivatives of  $u$  and  $v$  in trajectorial direction, we obtain the regularizer:

$$\mathcal{E}_t^{\text{2nd}} = \sum_{i=2}^3 \Psi_t \left( (u_{i+1} - 2u_i + u_{i-1})^2 + (v_{i+1} - 2v_i + v_{i-1})^2 \right). \quad (18)$$

In order to allow for motion discontinuities in trajectorial direction, we again apply the Charbonnier penalizer (15) with a contrast parameter  $\lambda_3 = 0.1$  in both cases. While the first-order regularizer  $\mathcal{E}_t^{\text{1st}}$  models the assumption of a piecewise smooth flow along the motion trajectory, the second-order regularizer  $\mathcal{E}_t^{\text{2nd}}$  actually allows piecewise linearly smooth

flow components in trajectory direction. Combining both regularizers yields

$$\mathcal{E}_t = \beta_1 \mathcal{E}_t^{1\text{st}} + \beta_2 \mathcal{E}_t^{2\text{nd}}, \quad (19)$$

where  $\beta_1 \geq 0$  and  $\beta_2 \geq 0$  are weights that balance the influence of the trajectory smoothness terms against the spatial smoothness term and the data term.

### 3.2. Adapting the Degree of Regularization

Since the aforementioned assumptions enter the final energy functional as a *soft constraint*, they are relaxed to a certain degree. However, it is clear that it should be avoided to impose such assumptions on sequences that do not fulfill the underlying trajectory smoothness model. Otherwise, an oversmoothing of the results can be expected. On the other hand, if it is known that the flow in a sequence obeys a certain trajectory motion model, enforcing exactly this model by an additional trajectory regularizer can be beneficial. As a consequence, we refrain from our first idea to combine first- and second-order models in a straightforward way as in (19). Instead we propose to *adapt the degree of trajectory regularization*, i.e. to choose either first-order, second-order, or no trajectory regularization. This adaptation can either be performed *globally* – one smoothness assumption is selected per image sequence – or *locally* – one smoothness assumption is selected per pixel, i.e. per trajectory. Similar adaptive schemes, but restricted to binary decisions only, became recently popular for globally enabling a geometry prior [21] or locally deciding between different data constancy assumptions [25]. In the following, we describe both our local and our global decision strategy.

In order to derive information about the degree of smoothness along the trajectories, we first compute the four optical flow fields  $\vec{w}_1$  to  $\vec{w}_4$  without a trajectory regularization. Then, for each location, we robustly fit a parabola to both the  $u$ - and  $v$ -component of the flow along the corresponding trajectory. This is done by minimizing

$$\mathcal{E}(a_u, b_u, c_u) = \int_{\Omega} \sum_{i=1}^4 \Psi_f \left( (a_u t_i^2 + b_u t_i + c_u - u_i)^2 \right) d\vec{x}, \quad (20)$$

$$\mathcal{E}(a_v, b_v, c_v) = \int_{\Omega} \sum_{i=1}^4 \Psi_f \left( (a_v t_i^2 + b_v t_i + c_v - v_i)^2 \right) d\vec{x}, \quad (21)$$

via iteratively reweighted least squares, where  $a_{u/v}(\vec{x})$ ,  $b_{u/v}(\vec{x})$  and  $c_{u/v}(\vec{x})$  are coefficient functions for  $u$  and  $v$ . The time instants  $t_i$  are chosen symmetrically w.r.t. the reference frame ( $t_1 = -1.5$ ,  $t_2 = -0.5$ ,  $t_3 = 0.5$  and  $t_4 = 1.5$ ) and  $\Psi_f$  is the Perona-Malik penalizer (14) with a contrast parameter  $\lambda_4 = 0.5$ .

Afterwards, we can infer the required degree of regularization from the computed coefficient functions for the quadratic and the linear term. Thus, we have to summarize

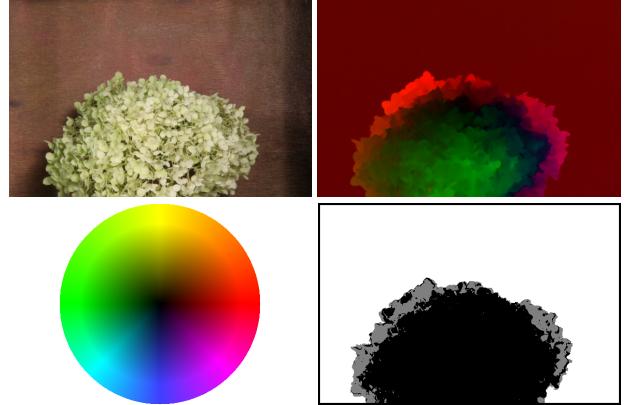


Figure 2. **Top row:** Reference frame (Hydrangea). Initial flow. **Bottom row:** Flow color code. Computed trajectory model (black: none, gray: 2nd order, white: 1st order).

the functions obtained by fitting to  $u$  and  $v$ . To this end, we propose the reduction

$$a(\vec{x}) = \max(|a_u|, |a_v|), \quad b(\vec{x}) = \max(|b_u|, |b_v|), \quad (22)$$

which strictly avoids oversmoothing by preferring larger coefficients and thus a higher order of regularization.

Now, in the case of our *locally adaptive* model, we determine the required degree of regularization based on two thresholds  $T_a$  and  $T_b$  using the following decision scheme:

Criterion	Trajectory model
$a(\vec{x}) > T_a$	none ( $\beta_1(\vec{x}) = \beta_2(\vec{x}) = 0$ )
$a(\vec{x}) \leq T_a \wedge b(\vec{x}) > T_b$	2nd order ( $\beta_1(\vec{x}) = 0$ )
$a(\vec{x}) \leq T_a \wedge b(\vec{x}) \leq T_b$	1st order ( $\beta_2(\vec{x}) = 0$ )

Here,  $\beta_1(\vec{x})$  and  $\beta_2(\vec{x})$  are the resulting local weighting functions for the first-order and second-order trajectory smoothness term in (19), respectively. An example for a computed trajectory model map is shown in Fig. 2.

In the case of our *globally adaptive* model, we first summarize the information in (22) by averaging the values of  $a(\vec{x})$  and  $b(\vec{x})$  over the image domain  $\Omega$ . Afterwards, we make a single decision for the global weights  $\beta_1$  and  $\beta_2$  on basis of the same assignment scheme as above.

## 4. Minimization

Combining the data term as well as the spatial and the trajectory smoothness terms into one energy, we obtain a non-convex functional that has to be minimized for the four flows. A common strategy to tackle this problem is to perform an incremental computation of the unknowns within a coarse-to-fine multi-scale approach. In this context, we follow the idea from [12] and approximate the original non-convex model by a series of convex energies in the flow increments with increasing resolution level. Similar to [20]

we can thus make use of a tensor notation which makes the convexity of the resulting energy functional explicit. Moreover, it allows us to apply a normalization strategy in the data term that makes deviations from the corresponding data constraints interpretable in a geometric way (cf. also [26]).

Minimizing the energy functional on each level w.r.t. the eight flow increments (four flows with two components each) then leads to the corresponding Euler-Lagrange equations. By discretizing them via finite difference approximations we obtain a system of equations on each level that is nonlinear due to the use of robust functions  $\Psi_d$ ,  $\Psi_{s_1}$ ,  $\Psi_{s_2}$  and  $\Psi_t$ . To ensure fast convergence, we solve this system with a bidirectional multigrid framework based on a nonlinear point coupled  $8 \times 8$  Gauß-Seidel solver that updates all unknowns in a pixel simultaneously [8]. In the coarse-to-fine multi-scale approach we use a downsampling factor of  $\eta = 0.95$ , while the images are warped onto the reference frame using Coons patches based on bicubic interpolation. On each level, three warps are performed. More details on the minimization can be found in [1].

## 5. Experiments

Let us now evaluate the performance of the proposed variational optical flow method based on temporal coherence (TC). To this end we computed results on the widely used Middlebury optical flow data sets [2]. We fixed the following parameters learned from the training data set by minimizing the average endpoint error (EPE) over all sequences:  $\alpha = 700$ ,  $\beta_1 = 90$ ,  $\beta_2 = 50$ ,  $\gamma = 20$ ,  $\sigma = 0.5$ ,  $\rho = 1.5$ . For the local adaptation scheme, we selected  $T_a = 0.028\mu$  and  $T_b = 0.014\mu$ , where  $\mu$  is the average magnitude of the four flow fields. For the global decision, the thresholds have been scaled down slightly by a factor 0.9.

In our first experiment we compare the performance of our TC approach against (i) the baseline method with *purely spatial* regularization (Complementary Optic Flow [26]), (ii) the baseline method with *spatio-temporal* regularization (also proposed in [26]) and (iii) a variant of the baseline method with *trajectorial stabilization* (following [3] but with stabilization forward and backward in time, see [1]). In all cases we used a RGB variant without automatic parameter estimation and optimized both the spatial and the temporal smoothness weights w.r.t. to the overall performance. The corresponding results are listed in the first three rows of Tab. 1. They clearly show that existing temporal regularization strategies fail in improving the quality of the estimation. While spatio-temporal methods are incapable of appropriately modeling motion trajectories with larger displacements, approaches based on trajectorial stabilization can only improve results when the assumption of a constant motion along the trajectory is fulfilled. Since this assumption does not hold for most of the training sequences, using a fixed temporal smoothness weight leads to consistently

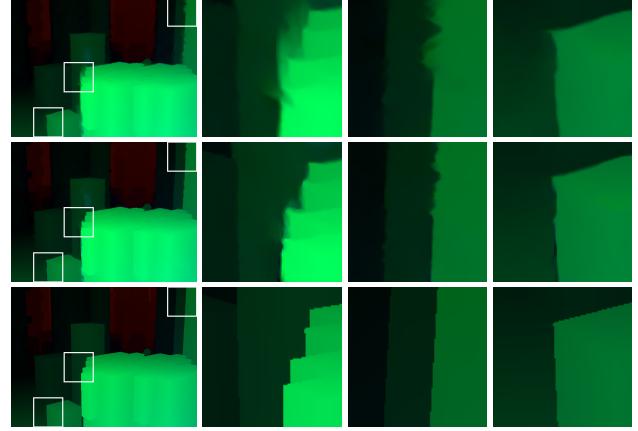


Figure 3. **Left to right:** Flow and three magnifications (Urban2). **Top to bottom:** Baseline method, TC method, ground truth.

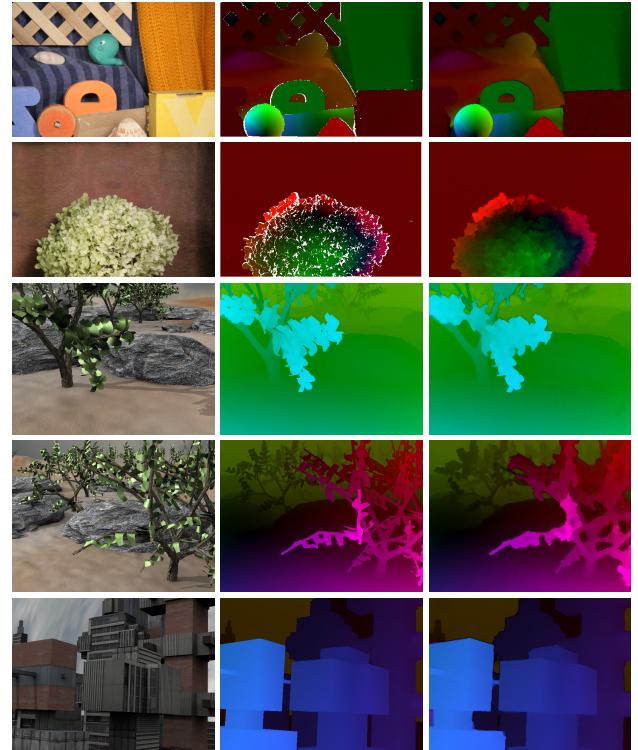


Figure 4. **Left to right:** Reference frame, ground truth, result. **Top to bottom:** RubberWhale, Hydrangea, Grove2, Grove3, Urban3.

worse results. In contrast, the results with our temporally coherent method (TC) that are given in the lower part of Tab. 1 show a significant quality gain. The overall improvement of our approach compared to the baseline method is also visually substantiated in Fig. 3. In this example, the sharpness of motion boundaries increases considerably.

In a second experiment, we further investigate the performance of our method by comparing several variants with different model components. Using the proposed trajectorial parametrization together with a *separate* spatial

Table 1. Comparison to the baseline method and alternative temporal regularization strategies from literature as well as evaluation of model components for proposed TC Flow method – Average Endpoint Error (EPE) for Middlebury *training* sequences with more than two images.

	Method	Avg.	RubberW.	Hydra.	Grove2	Grove3	Urban2	Urban3
Baseline method	COF spatial [26]	0.271	0.082	0.150	0.160	0.607	0.262	0.363
Alternative strategies	COF spatio-temporal [26]	0.491	0.165	0.333	0.246	0.794	0.622	0.787
	COF trajectoryl stabilization	0.313	0.088	0.153	0.166	0.634	0.288	0.550
Model components	TC separate spatial	0.261	0.076	0.146	0.140	0.600	0.192	0.409
	TC joint spatial	0.222	0.071	<b>0.134</b>	0.121	<b>0.529</b>	0.187	0.290
	TC 1st order trajectoryl	0.243	0.084	0.183	<b>0.114</b>	0.542	0.288	0.249
	TC 2nd order trajectoryl	0.227	0.086	0.182	<b>0.114</b>	0.540	0.203	<b>0.238</b>
	TC locally adaptive	0.219	<b>0.070</b>	0.141	0.115	0.541	<b>0.184</b>	0.261
	TC globally adaptive	<b>0.214</b>	0.071	<b>0.134</b>	<b>0.114</b>	0.540	0.187	<b>0.238</b>

Average endpoint error	avg. rank	Army (Hidden texture)			Mequin (Hidden texture)			Schefflera (Hidden texture)			Wooden (Hidden texture)			Grove (Synthetic)			Urban (Synthetic)			Yosemite (Synthetic)			Teddy (Stereo)						
		GT	im0	im1	all	disc	untext	GT	im0	im1	all	disc	untext	GT	im0	im1	all	disc	untext	GT	im0	im1	all	disc	untext	GT	im0	im1	all
MDP-Flow2 [40]	4.7	<b>0.096</b>	0.233	0.073	<b>0.162</b>	<b>0.521</b>	0.133	<b>0.222</b>	0.463	0.174	<b>0.17</b>	0.9316	0.094	<b>0.654</b>	<b>0.988</b>	<b>0.434</b>	<b>0.291</b>	<b>0.911</b>	<b>0.263</b>	<b>0.116</b>	<b>0.137</b>	<b>0.177</b>	<b>0.516</b>	<b>1.117</b>	<b>0.726</b>				
Layers++ [38]	5.5	<b>0.082</b>	0.211	0.079	<b>0.196</b>	0.563	0.173	<b>0.201</b>	<b>0.401</b>	0.188	<b>0.13</b>	<b>0.581</b>	<b>0.071</b>	<b>0.481</b>	<b>0.701</b>	<b>0.333</b>	<b>0.4710</b>	<b>1.012</b>	<b>0.333</b>	<b>0.1522</b>	<b>0.1416</b>	<b>0.2421</b>	<b>0.461</b>	<b>0.881</b>	<b>0.726</b>				
TC-Flow [48]	7.5	<b>0.071</b>	<b>0.211</b>	<b>0.066</b>	<b>0.151</b>	<b>0.594</b>	<b>0.111</b>	<b>0.3112</b>	<b>0.7819</b>	<b>0.1411</b>	<b>0.164</b>	<b>0.866</b>	<b>0.082</b>	<b>0.756</b>	<b>1.114</b>	<b>0.549</b>	<b>0.426</b>	<b>1.4014</b>	<b>0.256</b>	<b>0.116</b>	<b>0.123</b>	<b>0.2936</b>	<b>0.6210</b>	<b>1.3511</b>	<b>0.9321</b>				
LSM [41]	8.1	<b>0.082</b>	0.233	0.073	<b>0.2213</b>	<b>0.7314</b>	<b>0.1816</b>	<b>0.288</b>	<b>0.646</b>	<b>0.1912</b>	<b>0.143</b>	<b>0.702</b>	<b>0.094</b>	<b>0.666</b>	<b>0.974</b>	<b>0.486</b>	<b>0.5012</b>	<b>1.064</b>	<b>0.339</b>	<b>0.1522</b>	<b>0.1233</b>	<b>0.2930</b>	<b>0.504</b>	<b>0.993</b>	<b>0.738</b>				
Classic+NL [31]	8.9	<b>0.082</b>	0.233	0.073	<b>0.2213</b>	<b>0.7416</b>	<b>0.1816</b>	<b>0.298</b>	<b>0.685</b>	<b>0.1912</b>	<b>0.1533</b>	<b>0.734</b>	<b>0.094</b>	<b>0.643</b>	<b>0.933</b>	<b>0.475</b>	<b>0.5214</b>	<b>1.126</b>	<b>0.339</b>	<b>0.1628</b>	<b>0.137</b>	<b>0.2930</b>	<b>0.493</b>	<b>0.982</b>	<b>0.749</b>				
MDP-Flow [26]	10.5	<b>0.096</b>	0.258	0.080	<b>0.196</b>	<b>0.542</b>	<b>0.1816</b>	<b>0.244</b>	<b>0.556</b>	<b>0.2016</b>	<b>0.166</b>	<b>0.9113</b>	<b>0.094</b>	<b>0.747</b>	<b>1.067</b>	<b>0.6112</b>	<b>0.469</b>	<b>1.023</b>	<b>0.3514</b>	<b>0.1210</b>	<b>0.1416</b>	<b>0.177</b>	<b>0.7524</b>	<b>1.6826</b>	<b>0.9723</b>				
OFH [39]	11.0	<b>0.1011</b>	<b>0.258</b>	<b>0.096</b>	<b>0.196</b>	<b>0.6910</b>	<b>0.146</b>	<b>0.4318</b>	<b>1.0221</b>	<b>0.174</b>	<b>0.1719</b>	<b>1.0821</b>	<b>0.082</b>	<b>0.8716</b>	<b>1.2514</b>	<b>0.7318</b>	<b>0.4368</b>	<b>1.6923</b>	<b>0.327</b>	<b>0.103</b>	<b>0.137</b>	<b>0.1810</b>	<b>0.599</b>	<b>1.4014</b>	<b>0.749</b>				
OF-Mol [49]	11.4	<b>0.082</b>	0.233	0.073	<b>0.2826</b>	<b>0.9929</b>	<b>0.2022</b>	<b>0.288</b>	<b>0.646</b>	<b>0.1912</b>	<b>0.166</b>	<b>0.806</b>	<b>0.094</b>	<b>0.759</b>	<b>1.1210</b>	<b>0.507</b>	<b>0.5214</b>	<b>1.095</b>	<b>0.339</b>	<b>0.1628</b>	<b>0.137</b>	<b>0.3033</b>	<b>0.568</b>	<b>1.086</b>	<b>0.7811</b>				
Complementary OF [21]	16.9	<b>0.1117</b>	0.2816	0.1023	<b>0.183</b>	<b>0.637</b>	<b>0.1222</b>	<b>0.3121</b>	<b>0.7512</b>	<b>0.188</b>	<b>0.1914</b>	<b>0.9717</b>	<b>0.1217</b>	<b>0.9725</b>	<b>1.3121</b>	<b>1.0029</b>	<b>1.7843</b>	<b>1.7324</b>	<b>0.8736</b>	<b>0.1116</b>	<b>0.123</b>	<b>0.2218</b>	<b>0.6815</b>	<b>1.4816</b>	<b>0.9522</b>				

Figure 5. Snapshot of the Middlebury optical flow benchmark at the time of submission (February 27th, 2011). The proposed method is called *TC-Flow* and the baseline method is *Complementary OF*. The run time on the Urban sequence ( $640 \times 480$ ) was 2500 seconds.

regularizer (10)–(11) gives slightly more accurate flow fields in almost all cases. This already shows that incorporating the information from several frames can be useful. A huge improvement of results can be observed when using a *joint* spatial regularizer (12)–(13). This can be attributed to the underlying assumption of temporally coherent spatial flow structures: While motion boundaries become spatially aligned along the trajectory, no temporal regularity of the flows is enforced. In cases where trajectoryl smoothness of flow fields holds, adding a first or second-order trajectoryl regularizer can further improve the quality, see e.g. Grove2 and Urban3. However, if such an assumption is violated, a strict enforcement of trajectoryl smoothness leads to a significant deterioration. It is thus not surprising that adaptive models give the best results in most cases. A visual comparison of flow fields obtained with the globally adaptive TC method to the ground truth can be found in Fig. 4.

In our final experiment, we compare the performance of the proposed globally adaptive TC method to the state-of-the-art by using the Middlebury optical flow benchmark (<http://vision.middlebury.edu/flow/eval>). For the Teddy sequence, where only two frames are available, our method essentially reduces to the baseline method. As one can see from the snapshot in Fig. 5, our approach ranks among the Top 3 methods and significantly outperforms the baseline technique “Complementary OF” [26] (average rank of 7.5 compared to 16.9). Particularly in untextured regions our method yields an unprecedented quality. Evidently, incorporating appropriate assumptions on temporal coherence pays off in terms of quality.

Further experimental results can be found in [1].

## 6. Conclusions and Outlook

We have addressed the problem of appropriately modeling temporal coherence in variational optical flow estimation from multiple images. In this context, we contributed four important concepts: (i) We parametrized all flow fields w.r.t. one reference frame and thus obtained a natural temporal representation along motion trajectories. (ii) This parametrization enabled us to formulate a spatial regularizer that encourages a temporally coherent flow structure. Our experiments have demonstrated that this assumption is generally useful as it aligns the structure of all flow fields without making any assumption on their temporal evolution. (iii) We additionally introduced trajectoryl regularizers of first and second order that further improved the results in such cases where our motion models were appropriate. In this context, it became obvious that an adaptive approach is mandatory as many sequences exhibit complicated motion patterns. (iv) Consequently, we proposed to adapt the degree of trajectoryl regularization either locally or globally. Despite of the already strong spatial coupling of the flows, we could still observe a further improvement of the results.

We hope that our work sparks more research on optical flow and related problems that can benefit from assumptions on temporal coherence. As future work we plan to investigate more sophisticated motion models and other ideas that have been established in the context of tracking.

### Acknowledgements.

This work has been partly funded by the Cluster of Excellence “Multimodal Computing and Interaction” within the Excellence Initiative of the German Federal Government.

## References

- [1] <http://vip.mmc.uni-saarland.de/research/iccv11/index.shtml>.
- [2] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, Sept. 2010.
- [3] M. J. Black and P. Anandan. Robust dynamic motion estimation over time. In *Proc. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 292–302, Maui, HI, June 1991. IEEE Computer Society Press.
- [4] M. J. Black and P. Anandan. The robust estimation of multiple motions: parametric and piecewise smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, Jan. 1996.
- [5] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In T. Pajdla and J. Matas, editors, *Computer Vision – ECCV 2004, Part IV*, volume 3024 of *Lecture Notes in Computer Science*, pages 25–36. Springer, Berlin, 2004.
- [6] T. Brox, A. Bruhn, and J. Weickert. Variational motion segmentation with level sets. In H. Bischof, A. Leonardis, and A. Pinz, editors, *Computer Vision – ECCV 2006, Part I*, volume 3951 of *Lecture Notes in Computer Science*, pages 471–483. Springer, Berlin, 2006.
- [7] A. Bruhn and J. Weickert. Towards ultimate motion estimation: Combining highest accuracy with real-time performance. In *Proc. Tenth International Conference on Computer Vision*, volume 1, pages 749–755, Beijing, China, Oct. 2005. IEEE Computer Society Press.
- [8] A. Bruhn, J. Weickert, T. Kohlberger, and C. Schnörr. A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. *International Journal of Computer Vision*, 70(3):257–277, Dec. 2006.
- [9] K. Chaudhury and R. Mehrotra. A trajectory-based computational model for optical flow estimation. *IEEE Transactions on Robotics and Automation*, 11(5):733–741, 1995.
- [10] T. M. Chin, C. Karl, and A. S. Willsky. Probabilistic and sequential computation of optical flow using temporal coherence. *IEEE Transactions on Image Processing*, 3(6):773–788, Nov. 1994.
- [11] M. Elad and A. Feuer. Recursive optical flow estimation – adaptive filtering approach. *Journal of Visual Communication and Image Representation*, 9(2):119–138, June 1998.
- [12] E. Mémin and P. Pérez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Transactions on Image Processing*, 7(5):703–719, May 1998.
- [13] D. W. Murray and B. F. Buxton. Scene segmentation from visual motion using global optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(2):220–228, Mar. 1987.
- [14] H.-H. Nagel. Extending the 'oriented smoothness constraint' into the temporal domain and the estimation of derivatives of optical flow. In O. Faugeras, editor, *Computer Vision – ECCV '90*, volume 427 of *Lecture Notes in Computer Science*, pages 139–148. Springer, Berlin, 1990.
- [15] A. Salgado and J. Sánchez. Temporal constraints in large optical flow estimation. In R. Moreno-Díaz, F. Pichler, and A. Quesada-Arencibia, editors, *Computer Aided Systems Theory – EUROCAST 2007*, volume 4739 of *Lecture Notes in Computer Science*, pages 709–716. Springer, Berlin, 2007.
- [16] P. Sand and S. Teller. Particle video: Long-range motion estimation using point trajectories. In *Proc. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 2195–2202, New York, NY, June 2006. IEEE Computer Society Press.
- [17] D. Sun, S. Roth, J. P. Lewis, and M. J. Black. Learning optical flow. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision – ECCV 2008, Part III*, volume 5304 of *Lecture Notes in Computer Science*, pages 83–97. Springer, Berlin, 2008.
- [18] D. Sun, E. Sudderth, and M. J. Black. Layered image motion with explicit occlusions, temporal consistency, and depth ordering. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 2226–2234. 2010.
- [19] N. Sundaram, T. Brox, and K. Keutzer. Dense point trajectories by GPU-accelerated large displacement optical flow. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Computer Vision – ECCV 2010*, volume 6311 of *Lecture Notes in Computer Science*, pages 438–451. Springer, Berlin, 2010.
- [20] L. Valgaerts, A. Bruhn, H. Zimmer, J. Weickert, C. Stoll, and C. Theobalt. Joint estimation of motion, structure and geometry from stereo sequences. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Computer Vision – ECCV 2010*, volume 6314 of *Lecture Notes in Computer Science*, pages 568–581. Springer, Berlin, 2010.
- [21] A. Wedel, D. Cremers, T. Pock, and H. Bischof. Structure-and motion-adaptive regularization for high accuracy optic flow. In *Proc. Twelfth International Conference on Computer Vision*, volume 1, pages 1663–1668, Kyoto, Japan, Sept. 2009. IEEE Computer Society Press.
- [22] J. Weickert and C. Schnörr. Variational optic flow computation with a spatio-temporal smoothness constraint. *Journal of Mathematical Imaging and Vision*, 14(3):245–255, May 2001.
- [23] M. Werlberger, T. Pock, and H. Bischof. Motion estimation with non-local total variation regularization. In *Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, June 2010. IEEE Computer Society Press.
- [24] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof. Anisotropic Huber- $L^1$  optical flow. In *Proc. 20th British Machine Vision Conference*, London, UK, Sept. 2009. British Machine Vision Association.
- [25] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. In *Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1293–1300, San Francisco, CA, June 2010. IEEE Computer Society Press.
- [26] H. Zimmer, A. Bruhn, and J. Weickert. Optic flow in harmony. *International Journal of Computer Vision*, 93(3):368–388, Apr. 2011.