



Variational Large Displacement Optical Flow Without Feature Matches

Michael Stoll^(✉), Daniel Maurer, and Andrés Bruhn

University of Stuttgart, 70569 Stuttgart, Germany
{stoll,maurer,bruhn}@vis.uni-stuttgart.de
<http://cvvis.visus.uni-stuttgart.de/index.shtml>

Abstract. The optical flow within a scene can be an arbitrarily complex composition of motion patterns that typically differ regarding their scale. Hence, using a single algorithm with a single set of parameters is often not sufficient to capture the variety of these motion patterns. In particular, the estimation of large displacements of small objects poses a problem. In order to cope with this problem, many recent methods estimate the optical flow by a fusion of flow candidates obtained either from different algorithms or from the same algorithm using different parameters. This, however, typically results in a pipeline of methods for estimating and fusing the candidate flows, each requiring an individual model with a dedicated solution strategy. In this paper, we investigate what results can be achieved with a pure variational approach based on a standard coarse-to-fine optimization. To this end, we propose a novel variational method for the simultaneous estimation and fusion of flow candidates. By jointly using multiple smoothness weights within a single energy functional, we are able to capture different motion patterns and hence to estimate large displacements even without additional feature matches. In the same functional, an intrinsic model-based fusion allows to integrate all these candidates into a single flow field, combining sufficiently smooth overall motion with locally large displacements. Experiments on large displacement sequences and the Sintel benchmark demonstrate the feasibility of our approach and show improved results compared to a single-smoothness baseline method.

1 Introduction

The estimation of optical flow has been a core problem in computer vision for decades. Many successful methods for solving this task belong to the class of variational approaches. Based on the minimization of a continuous energy functional consisting of a data and a smoothness term, such methods offer dense and sub-pixel accurate results as well as a transparent modelling. Since the pioneering work of Horn and Schunck [14], a lot of progress has been made on both the modelling and the optimization side. On the modelling side, modern smoothness priors allow the estimation of flow fields with both gradual transitions [6, 11] and sharp motion discontinuities [16, 27], while modern data terms cope with noise [4], outliers and varying illumination [11]. On the optimization side, coarse-to-fine schemes [17] have been proposed that allow to handle large displacements

of relatively large objects. Fast motions of small objects, however, are still hard to capture.

One widely used approach to alleviate this problem is the integration of descriptor matches [8, 21, 25]. While such methods allow to handle arbitrarily large motion, they heavily rely on the uniqueness of the underlying descriptors. Hence, in case of weakly textured regions or repetitive patterns, such methods are likely to produce false matches which deteriorate the final optical flow estimation. Recent approaches face this problem by applying a-posteriori regularization to a set of descriptor matches in order to improve its quality [12].

While there are scenarios where the large displacement problem of small objects is intrinsically unsolvable – e.g. in the presence of multiple non-unique instances – a surprisingly large share of large displacement cases can actually be solved even *with a-priori* regularization, i.e. regularization during the estimation of the matches. In order to understand in which cases large displacements can still be recovered correctly, we have to distinguish two scenarios: (i) One problem is that small objects may not be present on that coarse-to-fine level that is necessary for the estimation of the displacement [8]. This case cannot be handled by standard coarse-to-fine optimization without further data transformations [18]. (ii) Another problem – which, however, has hardly been addressed in the literature so far – is the influence of the balance between data and smoothness term on the estimation of large displacements. For small objects that undergo large displacements, it is typically cheaper to violate the constancy assumptions in the data term (due to the small spatial extent) than to violate the regularity assumptions in the smoothness term (due to the large motion gradient). This particularly holds for large values of the smoothness parameter that are typically required to obtain noise-free flow fields. So even if there is sufficient data on the appropriate coarse-to-fine level, the smoothness term will suppress the estimation of the corresponding large displacement.

In this context, Brox and Malik [8] made the observation that fast motion of high-contrast objects is more likely to be accurately estimated than the motion of low-contrast objects. This is related to the fact that there is an implicit weighting of the constancy assumptions with the corresponding image gradient as observed in [27]. In view of the data costs, mismatches of high-contrast objects are thus more expensive than those of low-contrast objects. This, in turn, suggests to use constraint normalization as in [27] when estimating large displacements.

Contributions. In this work, we address the aforementioned problem of that the appropriate smoothness weight may depend on the local motion pattern. By proposing a variational method that jointly estimates and fuses candidate flows with different smoothness weights into a final flow field, we show that many large displacement scenarios can actually be resolved without using additional feature matches. In contrast to related work from the literature that typically relies on a one-way pipeline based on a discrete fusion of pre-computed flows, we model the entire approach as a single minimization problem based on standard coarse-to-fine optimization. Moreover, we demonstrate the benefit of constraint normalization when estimating large displacements. Please note that we do not

focus on designing an overall top-performing method but rather on pushing the limits of pure variational approaches w.r.t. large displacements.

Related Work. To handle large displacements, Brox and Malik [8] proposed to integrate descriptor matches into variational methods by means of a similarity term. While Stoll *et al.* [21] improved the sensitivity of this strategy w.r.t. to outliers by restricting the integration of such matches to promising locations, Weinzaepfel *et al.* [25] investigated the use of improved descriptors. In contrast, Xu *et al.* [26] refrained from using a similarity term, and proposed to enhance the upsampled flow initialization by integrating SIFT-matches at each level of the coarse-to-fine optimization. In contrast to our work, all these methods rely on feature descriptors to estimate large displacements.

Tu *et al.* [23] used a similar strategy as [26] but they considered proposals generated by PatchMatch [3] and by varying the smoothness weight of a variational method. Similarly, Lempitsky *et al.* [15] considered flows obtained by different methods and different parameter sets in a discrete fusion approach. Both works [15, 23], however, did not investigate the benefit of varying the smoothness weight for large displacement optical flow.

In all cases, descriptor matching and match integration are separate steps.

2 Baseline Method

Let us start by introducing our baseline optical flow method which is the Complementary Optic Flow method [27]. It is a variational approach where the optical flow $\mathbf{w} = (u, v)^\top$ between two input color images $\mathbf{f}_1 = (f_1^1, f_1^2, f_1^3)^\top$ and $\mathbf{f}_2 = (f_2^1, f_2^2, f_2^3)^\top$ is computed as the minimizer of the following energy:

$$\mathcal{E}_{base}(\mathbf{w})_\alpha = \int_{\Omega} \mathcal{E}_D(\mathbf{w}) + \alpha \mathcal{E}_S(\mathbf{w}) \, dx. \quad (1)$$

Here, \mathcal{E}_D is the data term, \mathcal{E}_S is the smoothness term, $\alpha > 0$ is a balancing weight and $\mathbf{x} = (x, y)^\top \in \Omega$ is the location within the image domain $\Omega \subset \mathbb{R}^2$.

Data Term. The data term relates the two input images via the optical flow and is given by [27]

$$\begin{aligned} \mathcal{E}_D(\mathbf{w}) = & \delta \Psi_D \left(\sum_{c=1}^3 \left(\sqrt{\theta^c} \cdot (f_2^c(\mathbf{x} + \mathbf{w}) - f_1^c(\mathbf{x})) \right)^2 \right) \\ & + \gamma \Psi_D \left(\sum_{c=1}^3 \left| \begin{pmatrix} \sqrt{\theta_x^c} & 0 \\ 0 & \sqrt{\theta_y^c} \end{pmatrix} \cdot (\nabla f_2^c(\mathbf{x} + \mathbf{w}) - \nabla f_1^c(\mathbf{x})) \right|^2 \right). \quad (2) \end{aligned}$$

It comprises the brightness constancy and the gradient constancy assumption in order to allow for illumination robust flow estimation [7]. Moreover, to reduce the influence of large gradients, constraint normalization [20] is applied via the weights $\theta^c = 1/(|\nabla f_2^c|^2 + \zeta^2)$ and $\theta_*^c = 1/(|\nabla f_{2,*}^c|^2 + \zeta^2)$ (with $*$ $\in \{x, y\}$), where ζ is a regularization parameter that prevents divisions by zero. Finally, both

assumptions are rendered robust under noise by applying a sub-quadratic penalizer [4] – here given by the Charbonnier function [10] $\Psi_D(s^2) = 2\epsilon_D^2 \sqrt{1 + s^2/\epsilon_D^2}$ with contrast parameter ϵ_D . The non-negative weights δ, γ serve as balancing factors.

Smoothness Term. As smoothness term, we consider the anisotropic complementary smoothness term [27]

$$\mathcal{E}_S(\mathbf{w}) = \sum_{i=1}^2 \Psi_{S_i} \left(|\mathcal{J}\mathbf{w} \cdot \mathbf{r}_i|^2 \right), \quad (3)$$

that penalizes the directional derivatives of the flow by projecting the Jacobian \mathcal{J} onto the local directions $\mathbf{r}_1, \mathbf{r}_2$ of maximum and minimum information contrast. In this context, the directions \mathbf{r}_1 and \mathbf{r}_2 are the eigenvectors of the so-called *regularization tensor* [27] which reads

$$R_\rho = K_\rho * \sum_{c=1}^3 \left[\delta \nabla f_1^c \nabla f_1^{c\top} + \gamma \left(\nabla f_{1x}^c \nabla f_{1x}^{c\top} + \nabla f_{1y}^c \nabla f_{1y}^{c\top} \right) \right], \quad (4)$$

where $*$ denotes convolution with a Gaussian K_ρ of standard deviation ρ .

Following [24], we apply the edge-enhancing Perona-Malik penalizer [5] given by $\Psi_S(s^2) = \epsilon_{S_1}^2 \log(1 + s^2/\epsilon_{S_1}^2)$ in \mathbf{r}_1 -direction and the edge-preserving Charbonnier penalizer [10] in \mathbf{r}_2 -direction; the former with contrast parameter ϵ_{S_1} and the latter with contrast parameter ϵ_{S_2} .

3 Joint Estimation and Fusion Model

After we have discussed the baseline method in the previous section, we are now in the position to describe our joint estimation and fusion model. Similar to methods from the literature that include descriptor matches [8, 21, 25], we want to estimate an optical flow \mathbf{w}_f using the baseline method \mathcal{E}_{base} and some similarity term \mathcal{E}_{sim} that feeds N candidate flows $\mathbf{w} = \{\mathbf{w}_1, \dots, \mathbf{w}_N\}$ from the candidate model \mathcal{E}_{cand} into the solution. To this end, we propose the joint variational model

$$\mathcal{E}(\mathbf{w}, \mathbf{w}_f) = \mathcal{E}_{base}(\mathbf{w}_f)_{\alpha_f} + \mathcal{E}_{sim}(\mathbf{w}, \mathbf{w}_f) + \mathcal{E}_{cand}(\mathbf{w}), \quad (5)$$

that consists of three terms. On the one hand, as baseline model, we use the approach from the previous section with smoothness weight α_f . On the other hand, as candidate model, we consider multiple instances of the baseline model $\mathcal{E}_{base}(\mathbf{w})_{\alpha_i}$ with different smoothness weights α_i that estimate the corresponding candidate optical flows \mathbf{w}_i . It is given by

$$\mathcal{E}_{cand}(\mathbf{w}) = \lambda_C \cdot \sum_{i=1}^N \mathcal{E}_{base}(\mathbf{w}_i)_{\alpha_i} \quad (6)$$

Due to the different smoothness weights, the single instances can capture different levels of motion details, i.e. displacement scales. Finally, in order to couple

the candidate flows \mathbf{w}_i and the final optical flow \mathbf{w}_f , we introduce a similarity term \mathcal{E}_C for each of these instances weighted by a parameter β_i . The combined similarity term reads

$$\mathcal{E}_{sim}(\mathbf{w}, \mathbf{w}_f) = \sum_{i=1}^N \beta_i \mathcal{E}_C(\mathbf{w}, \mathbf{w}_f)_i, \quad (7)$$

where the distinct similarity terms are defined as

$$\mathcal{E}_C(\mathbf{w}, \mathbf{w}_f)_i = \int_{\Omega} c_i(\mathbf{x}, \mathbf{w}) \cdot \Psi_C (|\mathbf{w}_i - \mathbf{w}_f|^2) d\mathbf{x}. \quad (8)$$

Here, c_i is a local confidence function for the candidate flow \mathbf{w}_i and Ψ_C is the Charbonnier penalizer [10] that makes the estimation more robust against outliers in the candidate flows. In Sect. 4, we will define appropriate confidence functions c_i that steer the local influence of each instance flow \mathbf{w}_i on the final flow \mathbf{w}_f . The overall weight λ_C balances $\mathcal{E}_{cand}(\mathbf{w})$ and $\mathcal{E}_{base}(\mathbf{w}_f)$ by steering the direction of information flow between the candidate flows and the final flow. The higher it is, the more remains the estimation of the candidates \mathbf{w} unaffected by the similarity term and the information only flows from \mathbf{w} to \mathbf{w}_f via \mathcal{E}_{sim} while backward information flow is suppressed.

4 Smoothness Weights and Confidence Functions

Since we desire candidate flows at different smoothness scales, the questions arise how to choose the global smoothness weights of these flows and how to locally decide which flow candidate is the most appropriate. Let us discuss these two issues in the following sections.

4.1 Smoothness Weights

First of all, we define a maximum smoothness weight α_1 which is intended to be appropriate at most locations. Moreover, we consider smoothness weights that are significantly smaller in order to be able to capture large displacement motions. Our choice for the smoothness weights α_i of the flow candidates \mathbf{w}_i is an exponential decrease w.r.t. α_1 :

$$\alpha_i := \frac{\alpha_1}{2^{i-1}}. \quad (9)$$

With this choice, we can cover a wide range of different smoothness scales with only a low number of candidate flows. By the example of the Tennis sequence [8] depicted in Fig. 1 (top row), one can see at which smoothness scale the different motion patterns appear. While the first, smoothest flow covers the background motion and the overall motion of the Tennis player smoothly, the second flow covers the motion of the racket and the arm well, the third flow covers the motion of the hand and the right foot while the fifth flow covers the motion of the ball.

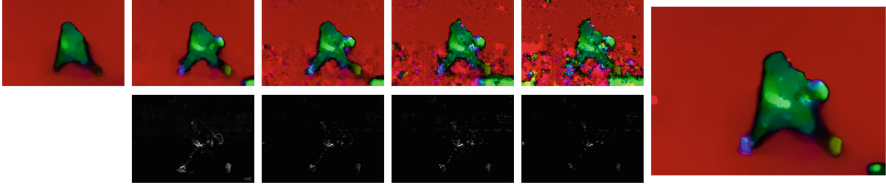


Fig. 1. Top row: Candidate flows with isotropic regularisation. **Bottom row:** normalized visualizations of the local confidence functions c_1, \dots, c_5 . **Right:** Final flow.

4.2 Assumptions on Local Confidences

Given a set of candidate flows \mathbf{w}_i with different smoothness scales, we take into account the considerations from the introduction to state the local assumptions on how to integrate these flows in the estimation of the final flow \mathbf{w}_f :

1. A less smooth flow is likely to fulfill the data term better than a smoother flow, independently from being reliable or unreliable. Hence, a less smooth flow shall only have influence if it provides significantly less data costs than both the next smoother flow candidate and the smoothest flow candidate (similar to considerations in [21]).
2. The less smooth a flow is, the more texture is necessary in order to achieve meaningful flow vectors (similar to [8]). Otherwise, we might likely get trapped into the aperture problem.
3. A less smooth flow should not be considered if the data is unreliable (i.e. in over- or undersaturated regions).

In order to integrate those assumptions in our local confidence functions c_i , we need measures for the data cost and for the local structure. While the data costs are simply given by evaluating the data term, we compute the structure tensor [13] to measure structureness [8], both on local patches to increase robustness.

4.3 Composition of the Local Confidence Function c_i

Following the assumptions from the last section, we model the local confidence function c_i (where i is the index of the candidate flow) as the product of three weights which will be defined in the following.

Structureness Weight. Let $s(\mathbf{x})$ be the smaller eigenvalue of the structure tensor (integrated over a 7×7 neighborhood) of the reference frame f_1 , let \bar{s} be its average value over the whole image and let $r_i = \frac{\alpha_1}{\alpha_i}$. The structureness weight is then defined as

$$w_i^s(\mathbf{x}) := \left(\frac{s(\mathbf{x})}{\bar{s}} \right)^{\kappa_s \cdot \log(r_i)}, \quad (10)$$

where the exponent κ_s is a free parameter. Here, the structureness weight is more pronounced for less smooth candidate flows (i.e. if r_i is bigger).

Cost Reduction Weight. Let \mathcal{E}_D be the data costs and let $\rho_{L \times L}(g, \mathbf{x})$ be a functional that averages the function g in a $L \times L$ neighborhood around \mathbf{x} . The following two functions describe the patch-wise energy improvement of flow \mathbf{w}_i compared to the previous, smoother flow \mathbf{w}_{i-1} and the first and smoothest flow \mathbf{w}_1 , respectively:

$$\begin{aligned} \delta_{\text{prev},L}(\mathbf{x}, \mathbf{w}, i) &:= \rho_{L \times L}(\mathcal{E}_D(\mathbf{w}_{i-1}), \mathbf{x}) - \rho_{L \times L}(\mathcal{E}_D(\mathbf{w}_i), \mathbf{x}), \\ \delta_{\text{first},L}(\mathbf{x}, \mathbf{w}, i) &:= \rho_{L \times L}(\mathcal{E}_D(\mathbf{w}_1), \mathbf{x}) - \rho_{L \times L}(\mathcal{E}_D(\mathbf{w}_i), \mathbf{x}). \end{aligned}$$

The cost reduction weight is then defined as

$$w_i^d(\mathbf{x}) := \log \left(1 + e^{\kappa_d(\delta_{\text{prev},L}(\mathbf{x}, \mathbf{w}, i) + \delta_{\text{first},L}(\mathbf{x}, \mathbf{w}, i))} \right), \quad (11)$$

where κ_d is a free parameter. Please note that this function resembles a linear one for large arguments of the exponential while it approaches zero for decreasing (negative) arguments.

Data Reliability Weight. We define $\chi_I(\mathbf{x})$ as an indicator function that excludes under- and oversaturated regions. It reads

$$\chi_I(\mathbf{x}) = \begin{cases} 1 & \text{if } f_1^c(\mathbf{x}) > \tau \text{ and } f_1^c(\mathbf{x}) < 255 - \tau \quad \forall c \in \{1, 2, 3\}, \\ 0 & \text{else} \end{cases}, \quad (12)$$

where $\tau = 1$ is a robustness threshold.

Overall Confidence Function. The overall confidence functions c_1, \dots, c_N are then defined as follows

$$\hat{c}_i(\mathbf{x}, \mathbf{w}) := w_i^d(\mathbf{x}) \cdot w_i^s(\mathbf{x}) \cdot \chi_I(\mathbf{x}) \quad (i > 1). \quad (13)$$

In order to be numerically robust, they are bounded from above via

$$c_i(\mathbf{x}, \mathbf{w}) := \min(\hat{c}_i(\mathbf{x}, \mathbf{w}), 1000). \quad (14)$$

Since the smoothest flow \mathbf{w}_1 serves as reference, it should be used everywhere except for those locations where a less smooth flow could improve the result. Hence, we define the confidence c_1 of the smoothest flow as

$$c_1(\mathbf{x}, \mathbf{w}) := 1, \quad (15)$$

which corresponds to the confidence of the other flows at average structured areas with only a small energy reduction.

Exemplary visualizations of these local confidence functions c_i for the Tennis sequence are shown in Fig. 1 (bottom row) where brighter values indicate higher confidence. As one can see, for each large displacement we have a high confidence in the smoothest candidate flow that is able to capture it.

5 Minimization

The whole variational model is minimized in a standard coarse-to-fine setting with warping and incremental computations [17]. Due to the nonlinearity of the penalizer functions, we additionally apply the lagged nonlinearity method in order to transform the nonlinear subproblems into series of linear equation systems. These linear equation systems are then solved using a multicolor variant of the successive overrelaxation (SOR) method [1].

Please note that in Eq. 8 the flow \mathbf{w} is apparent in both the confidence functions and the coupling term. In order to avoid multiplications of unknowns during the minimization, in each coarse-to-fine level we compute the confidence functions based on the flow from the previous level. This can also be seen as a lagged nonlinearity method regarding the computation of the confidences.

6 Evaluation

In order to evaluate the performance of our method, we conducted several experiments. These include a qualitative comparison against LDOF [8] that investigates the large displacement capabilities of our method, an experiment that analyzes the effect of constraint normalization in this context, an experiment that evaluates the effect of different types of data costs and a quantitative experiment on the MPI Sintel benchmark [9] that shows improvements compared to the baseline method. In all experiments, we optimized only the following parameters: the number N of candidates, the data weights δ and γ and the smoothness weight α_1 . To this end, we used the downhill simplex method as implemented in [22]. The remaining parameters are kept fixed throughout all experiments. They are given by $\beta_i = \alpha_f = \alpha_1$, $L = 5$, $\lambda_C = 1000$, $\kappa_s = 0.3$, $\kappa_d = 5$, $\epsilon_D = 0.01$, $\zeta = 0.01$, $\epsilon_{S_1} = 0.02$, $\epsilon_{S_2} = 0.03$.

6.1 Large Displacement Sequences

In our first experiment, we evaluate the performance of our method in the context of large displacements. To this end, we consider various challenging large displacement sequences from the literature and compare our results to those of the method of Brox and Malik (LDOF) [8] which has introduced descriptor matching in variational methods for large displacement optical flow. The parameters for all sequences are $\delta = \gamma = 0.5$, $\alpha_1 = 2$ and $N = 7$ candidate flows.

In Figs. 2 and 3 we show the results of both the publicly available implementation of LDOF and our novel variational method for large displacement optical flow. As one can see, our method correctly estimates the large displacements that LDOF is able to estimate – and even some more (see e.g. Tennis sequence 496). This particularly includes the displacements of the tennis balls that evidently extent their sizes. The extremely challenging Bird sequence [26] shows the limitations of both methods as none of them could capture the motion of the bird’s head. In order to demonstrate that the correct estimation of large displacements

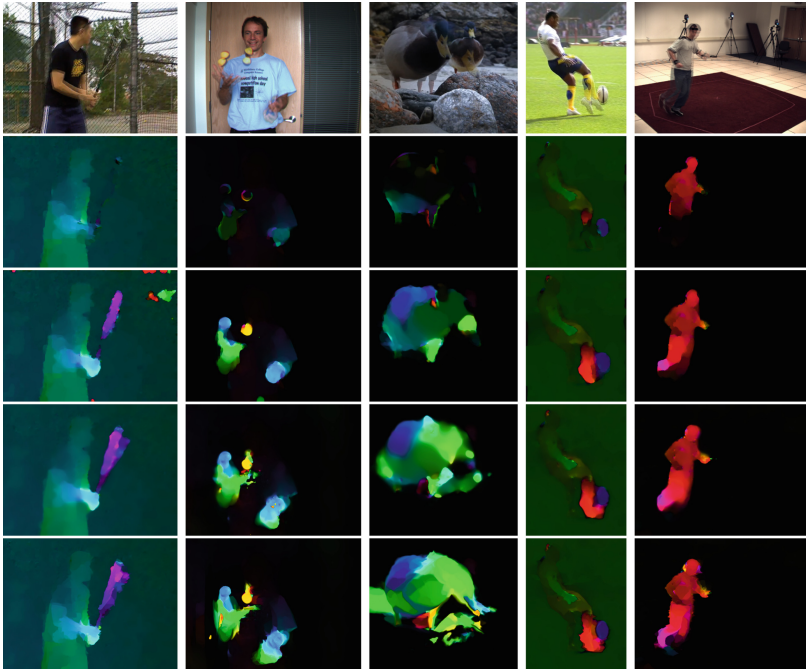


Fig. 2. Left to right: Baseball sequence [26], Beanbags sequence [2], Bird sequence, Football sequence [26], Human Eva sequence [19]. **Top to down:** Overlaid frames, baseline result, LDOF, our result (isotropic), our result (anisotropic).

does not depend on the anisotropic regularizer, we also added results for our method with an isotropic smoothness term (which is also used in LDOF).

While we have chosen the number of candidate flows fixed for all sequences, one may actually improve the results further by choosing it according to the extent of large displacements. For the beanbags sequences, already a value of $N = 3$ is sufficient, while we need a value of $N = 7$ in order to capture the motion of the tennis ball in the Tennis sequence 577.

6.2 Constraint Normalization

In our second experiment we show that constraint normalization [27] is helpful in the context of large displacements. To this end, we estimated flow fields without normalization and with normalization for different values of the normalization parameter ζ . While the general benefits of the constraint normalization have already been shown in [27], Fig. 4 shows the results on two large displacement sequences. As one can see particularly at hand of the tennis balls, both the deactivation of the constraint normalization and a too high value of ζ inhibit the estimation of large displacements. A too low value for ζ , in contrast, leads to noisier results. Using constraint normalization with a value between 0.001 and 0.01 (our standard value) for ζ provides the best results.

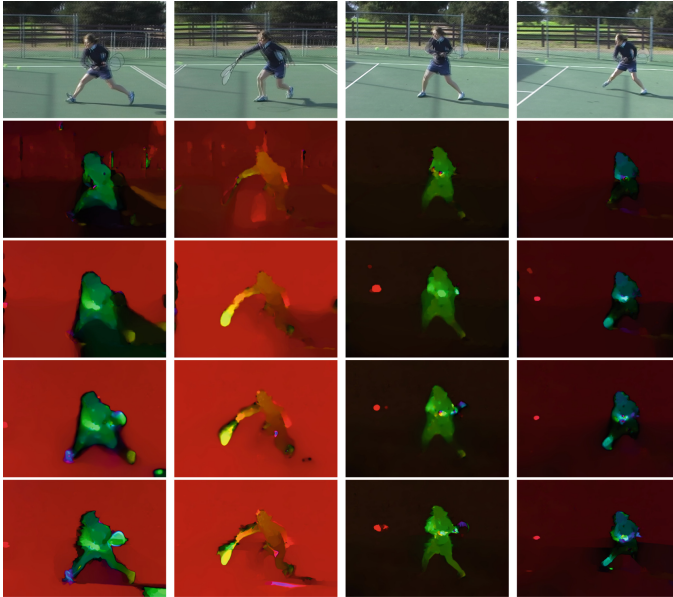


Fig. 3. Left to right: Tennis sequences 496, 502, 538, 577 [8]. **Top to down:** Overlaid frames, baseline result, LDOF, our result (isotropic), our result (anisotropic).

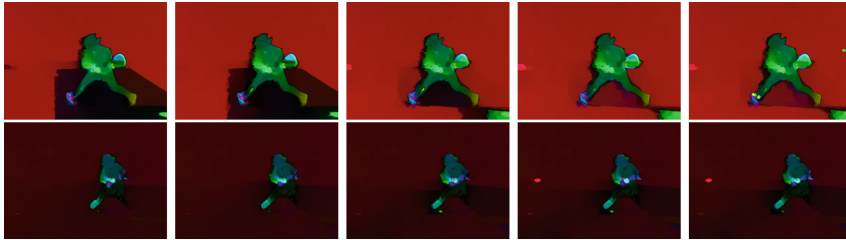


Fig. 4. From left to right: No constraint normalization, $\zeta = 1$, $\zeta = 0.1$, $\zeta = 0.001$, $\zeta = 0.00001$. **From top to bottom:** Tennis sequences 496 and 577.

6.3 Influence of the Data Constancy Assumptions

In our third experiment, we analyze the two types of data terms we used in our model w.r.t. their data costs and their influence on the fusion scheme. While the Brightness Constancy Assumption (BCA) can produce high costs at any part of a mismatched object, the Gradient Constancy Assumption (GCA) can only produce data costs where edges are involved. It is hence a lot sparser (see Fig. 5, top row). As can be seen from the bottom row of Fig. 5, the fusion using only the GCA data term is by far inferior to the results of using BCA or combining both data terms. The data costs of a pure GCA data term for incorrect matches are too low and hence it cannot compete with the smoothness term which prevents

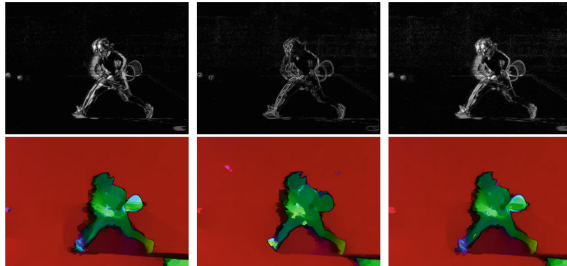


Fig. 5. From left to right: Brightness Constancy Assumption (BCA), Gradient Constancy Assumption (GCA) and both combined. **From top to bottom:** Data costs of the baseline flow (brighter grey values indicate larger energies), final result.

the motion discontinuity of a large displacement. In contrast, when including the BCA, the denser data costs make the misestimation of large displacements more expensive and thus increase the probability to estimate large displacements correctly. This shows that data costs with dense coverage for mismatched objects are important for our fusion scheme.

6.4 MPI Sintel Benchmark

In our fourth experiment, we compare our strategy with the baseline method (Complementary Optical Flow [27]) on the MPI Sintel benchmark [9]. To this end, we use our method with the first order complementary regularizer and computed results both for the training and the evaluation data.

Regarding the training data, Table 1 shows a clear improvement over the baseline ($N = 0$). The average endpoint error (AEE) decreases from 4.273 down to 3.974 (by 7%). This behavior is confirmed by the results for the evaluation data sets that are listed on the MPI Sintel webpage where our method is denoted as *ContFusion* and the baseline is denoted as *COF*. Here, the error decreases from 6.496 to 6.263 (by 3.6%) for the clean pass and from 8.204 to 7.857 (by 4.2%) for the final pass. This shows that our novel strategy of simultaneous estimation and fusion of motion candidates is also beneficial in a quantitative sense.

Table 1. Quantitative results on the clean training data of the MPI Sintel benchmark.

N	0	1	2	3	4	5	6
AEE	4.273	4.191	4.136	3.974	3.984	4.134	4.316

6.5 Limitations

The behavior at occlusions is a limitation of our method. This can be seen both visually at the large displacement sequences (in Figs. 2 and 3) and quantitatively

at the unmatched EPE in the MPI Sintel benchmark (that increases compared to the baseline). Additionally to regions with mismatched objects, occluded regions produce potentially high data costs. Since our confidence function heavily relies on data costs, correct smooth flows are replaced by less smooth candidate flows that lead to a smaller local data energy but are often meaningless.

7 Conclusion

In this work, we pushed the limits of variational approaches that are minimized using a standard coarse-to-fine scheme a little bit further w.r.t. large displacements. We have shown that many large displacement cases from the literature can be estimated without the need for descriptor matches. The weaknesses of prior variational methods in these cases are not due to weak data representations on coarse resolutions but due to a weight balancing of data term and smoothness term that is inappropriate for large displacement optical flow estimation. With multiple instances of the baseline model and appropriate choices of weighted similarity terms, we can estimate different scales of motions within a single variational model that simultaneously estimates and fuses candidate flows with different smoothness weights. The findings were confirmed by the evaluation which showed a good performance for large displacements and an improvement over its baseline method.

Limitations include the behavior at occluded regions where advanced occlusion handling would be necessary. Future work includes the handling of severe illumination changes where the BCA is not applicable at all and the GCA alone cannot help to estimate large displacements correctly, as well as the inclusion of second order smoothness terms for non-fronto-parallel motion patterns.

Acknowledgements. We thank the German Research Foundation (DFG) for financial support within project B04 of SFB/Transregio 161.

References

1. Adams, L., Ortega, J.: A multi-color sor method for parallel computation. In: Proceedings of International Conference on Parallel Processing, pp. 53–56 (1982)
2. Baker, S., Scharstein, D., Lewis, J.P., Roth, S., Black, M.J., Szeliski, R.: A database and evaluation methodology for optical flow. *Int. J. Comput. Vis.* **92**(1), 1–31 (2010)
3. Barnes, C., Shechtman, E., Goldman, D.B., Finkelstein, A.: The generalized Patch-Match correspondence algorithm. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6313, pp. 29–43. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15558-1_3
4. Black, M.J., Anandan, P.: Robust dynamic motion estimation over time. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 292–302 (1991)
5. Black, M.J., Anandan, P.: The robust estimation of multiple motions: parametric and piecewise smooth flow fields. *Comput. Vis. Image Underst.* **63**(1), 75–104 (1996)

6. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. *SIAM J. Imaging Sci.* **3**(3), 492–526 (2010)
7. Brox, T., Bruhn, A., Papenber, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: Pajdla, T., Matas, J. (eds.) *ECCV 2004*. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24673-2_3
8. Brox, T., Malik, J.: Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(3), 500–513 (2011)
9. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) *ECCV 2012*. LNCS, vol. 7577, pp. 611–625. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33783-3_44
10. Charbonnier, P., Blanc-Féraud, L., Aubert, G., Barlaud, M.: Two deterministic half-quadratic regularization algorithms for computed imaging. In: *Proceedings of IEEE International Conference on Image Processing*, pp. 168–172 (1994)
11. Demetz, O., Stoll, M., Volz, S., Weickert, J., Bruhn, A.: Learning brightness transfer functions for the joint recovery of illumination changes and optical flow. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014*. LNCS, vol. 8689, pp. 455–471. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_30
12. Drayer, B., Brox, T.: Combinatorial regularization of descriptor matching for optical flow estimation. In: *Proceedings of British Machine Vision Conference*, pp. 42.1–42.12 (2015)
13. Förstner, W., Gülch, E.: A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: *Proceedings of ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pp. 281–305 (1987)
14. Horn, B., Schunck, B.: Determining optical flow. *Artif. Intell.* **17**, 185–203 (1981)
15. Lempitsky, V., Roth, S., Rother, C.: FusionFlow: discrete-continuous optimization for optical flow estimation. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (2008)
16. Nagel, H.H., Enkelmann, W.: An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 565–593 (1986)
17. Papenber, N., Bruhn, A., Brox, T., Didas, S., Weickert, J.: Highly accurate optic flow computation with theoretically justified warping. *Int. J. Comput. Vis.* **67**(2), 141–158 (2006)
18. Sevilla-Lara, L., Sun, D., Learned-Miller, E.G., Black, M.J.: Optical flow estimation with channel constancy. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014*. LNCS, vol. 8689, pp. 423–438. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_28
19. Sigal, L., Balan, A.O., Black, M.J.: HumanEva: synchronized video and motion capture dataset for evaluation of articulated human motion. *Int. J. Comput. Vis.* **87**(1–2), 4–27 (2010)
20. Simoncelli, E.P., Adelson, E.H., Heeger, D.J.: Probability distributions of optical flow. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 310–315 (1991)

21. Stoll, M., Volz, S., Bruhn, A.: Adaptive integration of feature matches into variational optical flow methods. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012. LNCS, vol. 7726, pp. 1–14. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37431-9_1
22. Stoll, M., Volz, S., Maurer, D., Bruhn, A.: A time-efficient optimisation framework for parameters of optical flow methods. In: Sharma, P., Bianchi, F.M. (eds.) SCIA 2017. LNCS, vol. 10269, pp. 41–53. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59126-1_4
23. Tu, Z., Poppe, R., Veltkamp, R.C.: Weighted local intensity fusion method for variational optical flow estimation. *Pattern Recogn.* **50**, 223–232 (2016)
24. Volz, S., Bruhn, A., Valgaerts, L., Zimmer, H.: Modeling temporal coherence for optical flow. In: Proceedings of International Conference on Computer Vision, pp. 1116–1123 (2011)
25. Weinzaepfel, P., Revaud, J., Harchaoui, Z., Schmid, C.: Deepflow: large displacement optical flow with deep matching. In: Proceedings of International Conference on Computer Vision, pp. 1385–1392 (2013)
26. Xu, L., Jia, J., Matsushita, Y.: Motion detail preserving optical flow estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**, 1744–1757 (2012)
27. Zimmer, H., Bruhn, A., Weickert, J., Valgaerts, L., Salgado, A., Rosenhahn, B., Seidel, H.-P.: Complementary optic flow. In: Cremers, D., Boykov, Y., Blake, A., Schmidt, F.R. (eds.) EMMCVPR 2009. LNCS, vol. 5681, pp. 207–220. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-03641-5_16