

# Order-Adaptive and Illumination-Aware Variational Optical Flow Refinement

Daniel Maurer  
maurer@vis.uni-stuttgart.de

Michael Stoll  
stoll@vis.uni-stuttgart.de

Andrés Bruhn  
bruhn@vis.uni-stuttgart.de

Institute for Visualization and  
Interactive Systems  
University of Stuttgart, Germany

---

## Abstract

Variational approaches form an inherent part of most state-of-the-art pipeline approaches for optical flow computation. As the final step of the pipeline, the aim is to refine an initial flow field typically obtained by inpainting non-dense matches in order to provide highly accurate results. In this paper, we take advantage of recent improvements in variational optical flow estimation to construct an advanced variational model for this final refinement step. By combining an illumination aware data term with an order adaptive smoothness term, we obtain a highly flexible model that is able to cope well with a broad variety of different scenarios. Moreover, we propose the use of an additional reduced coarse-to-fine scheme instead of an exclusive initialisation scheme, which not only allows to refine the initialisation but also allows to correct larger erroneous displacements. Experiments on recent optical flow benchmarks show the advantages of the advanced variational refinement and the reduced coarse to fine scheme. The proposed order-adaptive method not only allows to significantly improve results compared to pipeline approaches based on traditional first-order refinement techniques, it also allows to outperform recent pure variational methods with full coarse-to-fine schemes.

## 1 Introduction

The estimation of motion information from image sequences is a key problem in computer vision and constitutes an important component for many high-level applications. Thereby, one is typically interested in determining the displacement vector field between two consecutive frames of an image sequence, the so-called optical flow. To solve this task, various techniques have been proposed in the literature. Among these, variational methods have a long and successful tradition, since they allow for both a transparent modelling as well as dense and accurate results [1]. They are based on the minimisation of an energy functional that comprises a data term and a smoothness term. While the data term imposes constancy assumptions on certain image features, the smoothness term regularises the often non-unique solution of the data term by imposing spatial regularity.

To make the optimisation feasible, variational methods often perform a linearisation of the highly non-convex data term. This, however, makes the estimation of large motion difficult, since the linearisation is typically only valid for small displacements. The standard

choice to overcome this shortcoming is the so-called warping strategy [6], which includes a coarse-to-fine estimation scheme. While this improves the estimation of large displacements, it does not resolve the problem for small objects, since they disappear on coarser scales. To cope with these large displacements of small objects, researchers proposed the integration of point correspondences, obtained via a preceding descriptor matching step [9, 29, 34] or obtained as an intermediate step to improve the initialisation at each coarse-to-fine level by additional candidate matches [32, 36]. Another approach is to replace the coarse-to-fine scheme by a suitable initialisation, obtained via a sparse-to-dense interpolation of point correspondences [28]. In fact, the latter approach is used in most state-of-the-art large displacement optical flow pipelines and is often referred to as *variational refinement* [0, 8, 11, 16, 23].

Since the variational refinement plays an essential role in many recent approaches, it is surprising that most of those methods rely on rather simple models for refinement. In particular, the refinement typically cannot keep up with the adaptivity and robustness of the preceding pipeline that consists of descriptor matching, filtering, and inpainting. Most prominent example is the widely used refinement model of the EpicFlow pipeline [34] that essentially combines a classical gradient constancy assumption with a simple isotropic first order smoothness term. In the last years, however, there has been significant progress in the modelling of variational methods. This progress includes more advanced data terms with a higher degree of invariance [9, 18, 24, 27], the joint estimation of motion and illumination changes [10], higher-order regularisers [3, 21, 26], as well as anisotropic [21, 37] and non-local smoothness terms [26, 35]. All those developments address important real-world problems such as varying illumination, the estimation of motion induced by a moving camera, or the sharp separation of motion boundaries. Hence, it is quite surprising that there have been no attempts in the literature so far to develop variational methods for optical flow refinement that consider these advanced concepts.

**Contributions.** In this work, we propose a novel model for variational refinement that combines robustness under varying illumination with the adaptive estimation of higher-order motion fields. While an illumination-aware data term is able to cope with locally affine illumination changes, an anisotropic order-adaptive smoothness term is able to produce solutions with gradual transitions where necessary while preserving sharp motion discontinuities at the same time. Moreover, we suggest a reduced coarse-to-fine scheme that is able to benefit from a good initialisation within the pipeline approach while still being able to correct errors in the intermediate results. The benefits of our method become explicit in the experimental evaluation. The experiments not only show improvements compared to conventional refinement schemes and pure variational methods, they also demonstrate good results on all major benchmarks such as KITTI 2012 [17], KITTI 2015 [22] and MPI Sintel [6].

**Related Work.** Conceptually closest related to our overall approach is the EpicFlow pipeline by Weinzaepfel *et al.* [34] as well as several follow-up works based on this pipeline. While most of these works focus on improving the matching step [0, 8, 11, 16, 23], there have hardly been any attempts to improve the sparse-to-dense inpainting [17] and, to the best of our knowledge, no attempts to improve the refinement step. Recent works based on discrete optimisation also make use of variational refinement [8, 23]. While these approaches do not necessarily suffer from the large displacement problem, they typically do not provide sub-pixel precise flow fields. Hence, they seek to gain additional precision by variational refinement. From a variational viewpoint, closest related to our work are the works of Demetz *et al.* [10] and Maurer *et al.* [20], which are based on traditional coarse-to-fine schemes. These methods, however, have difficulties to deal with fine structures and large displacements.

## 2 Pipeline for Large Displacement Optical Flow

Most state-of-the-art methods for large displacement optical flow use a pipeline approach as presented in [28]. Figure 1 illustrates the four main steps of the pipeline: matching, outlier filtering, inpainting and variational refinement. In the following we will detail on each of these four steps as they also form the basis of our algorithm.

**Matching.** The goal of the first step in the pipeline is the generation of input matches. Generally all kinds of different algorithms can be used, but as a matter of course the matches should be rather dense, so that a reasonable initialisation may be obtained. In our work we consider three different approaches to obtain input matches, whereof all are tailored to the problem of optical flow estimation. Our first choice is the *Deep Matching* approach [54], which creates matches by computing similarities of non-rigid patches. It is the favoured choice in the work [28]. Our second choice is the recent *CPM* method [16] – a coarse-to-fine variant of Patchmatch [9] – which basically is an approximate nearest neighbour field algorithm with an implicit regularisation. To measure the similarities of matches the CPM algorithm makes use of SIFT features [19]. Our last choice is *DiscreteFlow* [23]. In contrast to the other two approaches it contains explicit regularisation. To obtain the matches it first extracts a set of suitable proposals and optimises a cost function via dynamic programming. Like the other approaches it makes use of robust feature descriptors, in this case DAISY [51].

**Outlier Filtering.** The computed matches from the first step typically contain a certain amount of outliers, which occur for example due to occluded or low textured image regions. Since such erroneous matches can deteriorate the estimation substantially, it is common to perform some sort of outlier filtering such as bidirectional consistency checking and/or removal of small isolated segments. In practice, this second step does not eliminate all outliers, but considerably reduces their amount. In our approach we stick to the filtering steps as proposed by the respective matching approaches [16, 23, 54].

**Inpainting.** After removing outliers, the resulting flow field is typically non-dense. However, since the last step of the pipeline – the variational refinement – requires a dense flow field for initialisation, the missing locations have to be inpainted. For this purpose we use the locally-weighted affine estimation as presented in [28]. The algorithm is based on a weighted least-squares fit, where the weights are determined using a geodesic distance based on the image edges which are assumed to be a superset of the motion boundaries.

**Variational Refinement.** The final step refines the inpainted flow field using a variational method. Typically, this step aims at obtaining sub-pixel precision while it additionally introduces some kind of regularisation. In our case, we make use of two variational models that will be explained in the following sections: the commonly used EpicFlow model [28] that serves as baseline in our evaluation and our novel order-adaptive illumination-aware model.

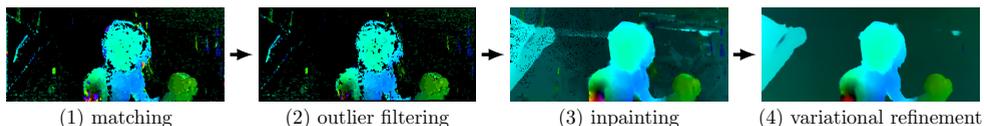


Figure 1: Illustration of the commonly used pipeline for large displacement optical flow by the example of a sequence from the MPI Sintel benchmark [6].

### 3 The EpicFlow Model

Let us start by discussing the commonly used EpicFlow refinement model [28]. To this end, let  $f(\mathbf{x}), g(\mathbf{x}) : \Omega \rightarrow \mathbb{R}^2$  denote two consecutive frames of an image sequence, where  $\mathbf{x} = (x, y)^\top$  denotes a position within the rectangular image domain  $\Omega \subset \mathbb{R}^2$ . Furthermore, let  $\mathbf{w} = (u, v)^\top : \Omega \rightarrow \mathbb{R}^2$  denote the optical flow. Then the EpicFlow model computes the refined flow as a minimiser of an energy functional of the form

$$E_{\text{epic}}(\mathbf{w}) = \int_{\Omega} D_{\text{epic}}(\mathbf{w}) + \alpha \cdot R_{\text{epic}}(\mathbf{w}) \, d\mathbf{x}, \quad (1)$$

where  $D_{\text{epic}}$  and  $R_{\text{epic}}$  denote the data term and smoothness term, respectively, and  $\alpha$  is the regularisation parameter that steers the relative impact of both terms.

**Data Term.** The data term  $D_{\text{epic}}$  comprises a brightness constancy assumption and a gradient constancy assumption with an additional constraint normalisation as proposed in [37]. After linearisation, the data term reads

$$D_{\text{epic}}(\mathbf{w}) = \Psi_{\text{d}} \left( \theta \cdot (f_x u + f_y v + f - g)^2 \right) + \lambda \cdot \Psi_{\text{d}} \left( \theta_x \cdot (f_{xx} u + f_{xy} v + f_x - g_x)^2 + \theta_y \cdot (f_{yx} u + f_{yy} v + f_y - g_y)^2 \right). \quad (2)$$

Here, subscripts of  $f$  and  $g$  denote partial derivatives,  $\lambda$  is a weighting parameter that balances the two constancy assumptions,  $\Psi_{\text{d}}(s^2) = \sqrt{s^2 + \varepsilon^2}$  with small  $\varepsilon > 0$  is a robust penaliser that implements the regularised absolute value function, and  $\theta, \theta_x, \theta_y$  are normalisation factors defined as

$$\theta = \frac{1}{|\nabla f|^2 + \zeta^2}, \quad \theta_x = \frac{1}{|\nabla f_x|^2 + \zeta^2}, \quad \theta_y = \frac{1}{|\nabla f_y|^2 + \zeta^2}, \quad (3)$$

where  $\nabla = (\partial_x, \partial_y)^\top$  is the spatial gradient and the parameter  $\zeta$  not only avoids a division by zero, but also reduces the influence of small gradients, e.g. noise in flat regions. While this data term works quite well in practice, it comes with the drawback that the gradient constancy assumption only allows to handle additive illumination changes but not multiplicative ones – in contrast to the descriptors of the initial matching process; see e.g. SIFT [19], DAISY [31].

**Smoothness Term.** In case of the smoothness term  $R_{\text{epic}}$  the EpicFlow model uses a first order flow-driven regulariser with image based weighting similar to [33], given by

$$R_{\text{epic}}(\mathbf{w}) = g(|\nabla f|) \cdot \Psi_{\text{s}} \left( |\nabla u|^2 + |\nabla v|^2 \right) \quad \text{with} \quad g(|\nabla f|) = \exp(-\kappa \cdot |\nabla f|), \quad (4)$$

where the robust penaliser function  $\Psi_{\text{s}}(s^2) = \Psi_{\text{d}}(s^2)$  allows to preserve motion discontinuities and the spatially adaptive weight  $g$  tries to align these discontinuities with image boundaries, i.e. it reduces the impact of the smoothness term at image edges depending on the parameter  $\kappa$ . This smoothness term has two major drawbacks: On the one hand, since it uses first order regularisation that prefers piecewise constant flow fields, it has problems with estimating highly non-fronto-parallel motion, e.g. affine motion that is typically present in ego motion scenes. On the other and, it does not make use of directional information to refine motion boundaries, which typically gives a less distinct separation of objects in the flow field compared to smoothness terms based on anisotropic regularisation.

## 4 Order-Adaptive Illumination-Aware Refinement

Based on the drawbacks of the EpicFlow model, we propose a novel order-adaptive and illumination-aware refinement model that combines recent concepts of variational optical flow estimation to eliminate these shortcomings. To this end, we consider the data term in [10], that explicitly models local illumination changes in terms of a set of coefficient fields  $\mathbf{c} = (c_1, \dots, c_N)^\top : \Omega \rightarrow \mathbb{R}^N$ , and the recent anisotropic order-adaptive regulariser in [24], that locally selects between first and second order regularisation using a spatially varying weighting function  $o : \Omega \rightarrow [0, 1]$ . Our novel energy has the following form

$$E_{\text{oir}}(\mathbf{w}, \mathbf{c}, o) = \int_{\Omega} D_{\text{illum}}(\mathbf{w}, \mathbf{c}) + \alpha \cdot R_{\text{oar}}(\mathbf{w}, o) + \beta \cdot R_{\text{illum}}(\mathbf{c}) + \gamma \cdot S_{\text{oar}}(o) \, d\mathbf{x}, \quad (5)$$

where the additional terms  $R_{\text{illum}}$  and  $S_{\text{oar}}$  are a regulariser for the illumination coefficients  $\mathbf{c}$  and a selection term for the order weighting function  $o$ , respectively. Let us now detail on the different components of the energy.

**Data Term.** As in the EpicFlow model the data term consists of a brightness and gradient constancy assumption. To account for more general illumination changes, however, we additionally make use of a parametrised brightness transfer function  $\Phi(g, \mathbf{c})$  in both assumptions. The resulting illumination-aware data term reads

$$D_{\text{illum}}(\mathbf{w}, \mathbf{c}) = \Psi_{\mathbf{c}} \left( \boldsymbol{\theta} \cdot (f_x u + f_y v + f - \Phi(g, \mathbf{c}))^2 \right) + \lambda \cdot \Psi_{\mathbf{c}} \left( \boldsymbol{\theta}_x \cdot (f_{xx} u + f_{xy} v + f_x - \partial_x \Phi(g, \mathbf{c}))^2 + \boldsymbol{\theta}_y \cdot (f_{yx} u + f_{yy} v + f_y - \partial_y \Phi(g, \mathbf{c}))^2 \right), \quad (6)$$

where  $\Psi_{\mathbf{c}}(s^2) = 2\varepsilon^2 \sqrt{1 + s^2/\varepsilon^2}$  is the robust Charbonnier penaliser [9] with contrast parameter  $\varepsilon > 0$ . While the general parametrised brightness transfer function [10, 13] is given by

$$\Phi(g, \mathbf{c}) = \bar{\phi}(g) + \sum_{i=1}^N c_i \cdot \phi_i(g), \quad (7)$$

where  $\phi_i(g) : \mathbb{R} \rightarrow \mathbb{R}$  denote the  $N$  basis functions and  $\bar{\phi}(g) : \mathbb{R} \rightarrow \mathbb{R}$  is the mean brightness transfer function, we choose  $\Phi(g, \mathbf{c})$  to be the normalised affine function, i.e.

$$\bar{\phi}(g) = g, \quad \phi_1(g) = \frac{g}{n_1}, \quad \phi_2(g) = \frac{1}{n_2}, \quad (8)$$

where  $n_1$  and  $n_2$  are normalisation factors such that  $\|\phi_i(g)\|_2 = 1$ . This choice of the brightness transfer function often constitutes a good trade-off between complexity and quality [10].

**Smoothness Term (Illumination).** In case of the smoothness term for the illumination coefficients we follow [10] and use a joint anisotropic first order regulariser which reads

$$R_{\text{illum}}(\mathbf{c}) = \sum_{m=1}^2 \Psi_m \left( \sum_{n=1}^N \left( \mathbf{r}_m^\top \nabla c_n \right)^2 \right). \quad (9)$$

It not only allows to adapt the regularisation locally to the underlying image structure in terms of two spatially varying directions  $\mathbf{r}_1$  and  $\mathbf{r}_2$  – these directions are obtained as eigenvectors of the regularisation tensor; see [10, 67]. It also allows to treat both directions differently which is reflected in the use of two separate penaliser functions  $\Psi_1(s^2)$  and  $\Psi_2(s^2)$

– in our case, we choose  $\Psi_1(s^2) = \varepsilon^2 \log(1 + s^2/\varepsilon^2)$  to be the Perona-Malik penaliser [25] and  $\Psi_2(s^2) = \Psi_c(s^2)$  to be the Charbonnier penaliser, both with contrast parameter  $\varepsilon > 0$ .

**Smoothness Term (Flow).** In case of the order-adaptive regulariser we choose the non-local selection scheme from [20], which has shown to perform equally well in scenes with fronto-parallel and affine motion. It is based on a combination of first and second order smoothness terms with locally varying weights. Its general form is given by

$$R_{\text{oar}}(\mathbf{w}, o) = \inf_{\mathbf{a}, \mathbf{b}} \left\{ \bar{o} \cdot S_1(\mathbf{w}) + (1 - \bar{o}) \cdot (S_2(\mathbf{w}, \mathbf{a}, \mathbf{b}) + T) + \delta \cdot S_3(\mathbf{a}, \mathbf{b}) \right\}, \quad (10)$$

where  $S_1$  is a first order regulariser and  $S_2$  and  $S_3$  form the coupling and smoothness term of a second order regulariser, respectively. In order to avoid over-fitting the data by only selecting the less restrictive second order regulariser, an activation cost  $T$  is introduced in the coupling term. Moreover, the selection process is rendered more robust by integrating the order weights  $o$  within a rectangular shaped neighbourhood  $\mathcal{N}(\mathbf{x})$  via

$$\bar{o}(\mathbf{x}) = \frac{1}{|\mathcal{N}(\mathbf{x})|} \int_{\mathcal{N}(\mathbf{x})} o(\mathbf{y}) d\mathbf{y}, \quad (11)$$

where  $|\mathcal{N}(\mathbf{x})|$  is the size of the neighbourhood.

Let us now detail on the employed first and second order smoothness terms. While the first order smoothness term is given by the anisotropic model [67]

$$S_1(\mathbf{w}) = \sum_{m=1}^2 \Psi_m \left( \left( \mathbf{r}_m^\top \nabla u \right)^2 + \left( \mathbf{r}_m^\top \nabla v \right)^2 \right), \quad (12)$$

the second order coupling approach is given by [14, 20]

$$S_2(\mathbf{w}, \mathbf{a}, \mathbf{b}) = \sum_{m=1}^2 \Psi_m \left( \left( \mathbf{r}_m^\top (\nabla u - \mathbf{a}) \right)^2 + \left( \mathbf{r}_m^\top (\nabla v - \mathbf{b}) \right)^2 \right), \quad (13)$$

which couples the flow gradients to the auxiliary functions  $\mathbf{a}$  and  $\mathbf{b}$  and

$$S_3(\mathbf{a}, \mathbf{b}) = \sum_{m=1}^2 \Psi_m \left( \sum_{l=1}^2 \left( \mathbf{r}_l^\top \mathcal{J} \mathbf{a} \mathbf{r}_m \right)^2 + \left( \mathbf{r}_l^\top \mathcal{J} \mathbf{b} \mathbf{r}_m \right)^2 \right), \quad (14)$$

that enforces smoothness on these auxiliary functions via penalising their Jacobians  $\mathcal{J} \mathbf{a}$  and  $\mathcal{J} \mathbf{b}$ . In this context, the weight  $\delta$  determines the amount of smoothness and both the directions  $\mathbf{r}_1, \mathbf{r}_2$  and the penaliser functions  $\Psi_1, \Psi_2$  are defined as before.

**Selection Term.** Finally, the selection term  $S_{\text{oar}}$  is given by

$$S_{\text{oar}}(o) = \ln(1 - o) - o \cdot \ln\left(\frac{1}{o} - 1\right), \quad (15)$$

which leads to an order adaptive selection via a sigmoid function based on the local energies of  $S_1$  and  $S_2$ . A detailed derivation of this selection term can be found in [20].

## 5 Minimisation

To minimise the energy functional (5) the well-known warping strategy is used, which is based on an incremental formulation of the unknowns with two nested fixed point iterations [5]. Thereby, the outer fixed point iteration is typically embedded in a coarse-to-fine scheme with downsampling factor  $\eta$  to avoid getting trapped in local minima of the energy. In our approach, however, in contrast to many pure variational methods [6, 10, 20, 26], we do not estimate the optical flow from scratch which would require to start at a very coarse resolution. Instead, we aim at refining the initial flow field provided by the preceding inpainting step such that we can benefit from a typically rather good initialisation. In this context, however, we also do not follow the other extreme, i.e. recent optic flow approaches that refine the initial flow field only at the finest resolution ( $\eta = 1$ ), to obtain sub-pixel precision [11, 8, 16, 23, 28]. Considering the problem that depending on the matching strategy, the initial matches can be off by several pixels – in particular if matching approaches operate with reduced image resolution, e.g. [8] – we propose a reduced coarse-to-fine scheme that starts the refinement at an intermediate level and hence still allows for sufficient corrections compared to a single scale refinement. For additional details we refer the reader to the supplementary material.

## 6 Evaluation

**Evaluation Setup.** The matching as well as the inpainting is performed using the publicly available code, provided by the respective authors. Thereby, all required parameters are set to the provided default values. In case of our proposed model we set most parameters fixed  $\lambda = 5$ ,  $\zeta = 10^{-2}$ ,  $\varepsilon = 10^{-2}$ ,  $T = 10^{-5}$ ,  $\gamma = 10^{-5}$  and only optimise the three remaining parameters  $\alpha$ ,  $\beta$ ,  $\delta$  using downhill simplex on the provided training data [30]. In order to avoid a bias towards a specific matching approach (DeepMatching, CPM, DiscreteFlow), we compute an individual set of these three parameters for each of the three approaches.

**Order-Adaptive Refinement.** In our first experiment we demonstrate the benefit of our order-adaptive refinement strategy compared to the first-order refinement of the EpicFlow model in case of highly non-fronto-parallel motion. Therefore, we depicted the results for a sequence of the KITTI 2012 benchmark [12] in Figure 2. Taking a look at the second row, which shows the inpainted CPM matches before variational refinement, one can see in the bad pixel (BP) visualisation that the affine inpainting did a good job at the bottom boundary, e.g. red framed region. When applying the refinement with the first order EpicFlow model afterwards this inpainted region deteriorates, but small displacements located at the image centre improve. In contrast, our order-adaptive refinement strategy (fourth row) improves both the inpainted areas as well as the small displacements located at the image centre. This finding is also reflected in the computed error measures of the entire KITTI 2012 benchmark for the CPM matches; see Table 1. While the BP error increases after the EpicFlow refinement from 10.99% to 14.58%, it decreases to 9.68% with our novel refinement.

**Reduced Coarse-to-fine Scheme.** In our second experiment we investigate the proposed reduced coarse-to-fine scheme. Using 10 resolution levels, we thereby compare three different settings which correspond to different initial scales:  $\eta = 1.0$  (no coarse-to-fine, i.e. full resolution),  $\eta = 0.95$  ( $0.63 \times$  full resolution) and  $\eta = 0.90$  ( $0.39 \times$  full resolution). The outcome is listed in Table 1. Here one can see, that in case of the KITTI benchmarks the results benefit quite a lot from the reduced coarse-to-fine scheme. In contrast, one cannot observe such an improvement for the MPI Sintel benchmark, but the results do not deteriorate either.

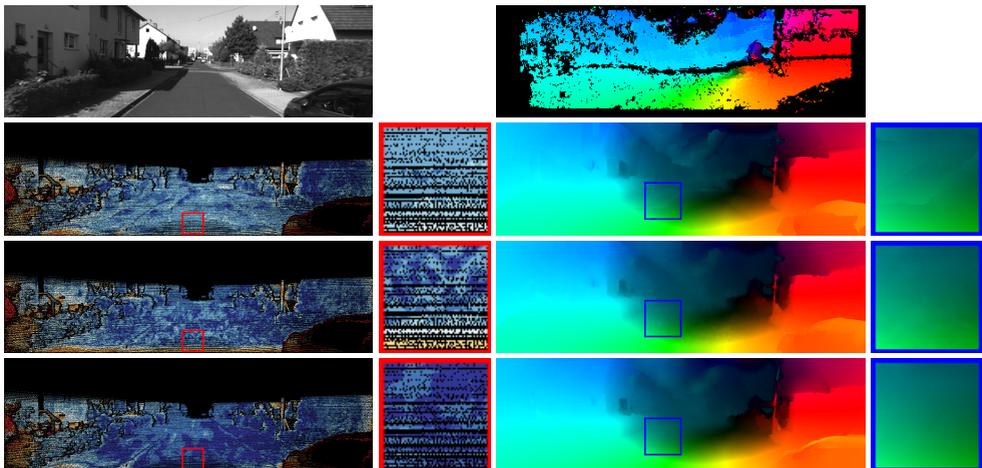


Figure 2: Training sequence #28 of the KITTI 2012 benchmark [10]. Best viewed electronically. **First row**: Reference image, CPM matches. **From left to right**: BP visualisation and flow field. **Second row**: Inpainted CPM matches (BP: 7.86%). **Third row**: EpicFlow refinement (BP: 10.80%). **Fourth row**: Proposed refinement (BP: 7.03%).

	KITTI 2012				KITTI 2015				Sintel
	noc		occ		noc		occ		clean
	AEE	BP (%)	AEE	BP (%)	AEE	BP (%)	AEE	BP (%)	AEE
<b>no refinement</b>									
Deepmatches	1.86	11.39	3.52	18.92	5.12	24.74	9.37	31.96	2.68
DiscreteFlow	<b>1.36</b>	7.25	3.06	16.02	<b>3.14</b>	<b>15.31</b>	<b>6.67</b>	24.37	2.21
CPM	1.43	<b>6.23</b>	<b>2.99</b>	<b>10.99</b>	3.66	16.43	7.77	<b>23.36</b>	<b>2.19</b>
<b>EpicFlow model</b>									
Deepmatches	1.42	7.64	3.24	16.24	4.71	20.06	9.18	28.38	2.27
DiscreteFlow	<b>1.17</b>	5.70	<b>2.97</b>	14.89	<b>2.94</b>	<b>13.62</b>	<b>6.68</b>	23.14	<b>1.94</b>
CPM	1.25	<b>5.37</b>	3.00	<b>14.58</b>	3.43	14.58	7.78	<b>22.86</b>	2.00
<b>our model (<math>\eta = 1.00</math>)</b>									
Deepmatches	1.32	7.65	2.83	12.82	4.59	19.60	8.82	26.14	2.26
DiscreteFlow	<b>1.08</b>	5.80	<b>2.54</b>	11.09	<b>2.83</b>	<b>13.03</b>	<b>6.29</b>	<b>20.08</b>	<b>1.91</b>
CPM	1.20	<b>5.66</b>	2.92	<b>10.19</b>	3.85	14.11	8.88	20.57	1.99
<b>our model (<math>\eta = 0.95</math>)</b>									
Deepmatches	1.20	6.28	2.61	10.91	4.45	17.69	8.45	23.89	2.24
DiscreteFlow	<b>1.02</b>	<b>5.05</b>	<b>2.39</b>	<b>9.77</b>	<b>2.79</b>	<b>12.43</b>	<b>5.99</b>	<b>18.56</b>	<b>1.91</b>
CPM	1.14	5.20	2.79	9.83	3.25	13.39	7.43	19.43	2.01
<b>our model (<math>\eta = 0.90</math>)</b>									
Deepmatches	1.16	5.67	2.52	10.06	4.32	16.25	8.25	22.33	2.23
DiscreteFlow	<b>1.01</b>	<b>4.87</b>	<b>2.34</b>	<b>9.29</b>	<b>2.77</b>	<b>12.16</b>	<b>5.89</b>	<b>18.10</b>	<b>1.94</b>
CPM	1.14	5.18	2.78	9.68	3.24	13.25	7.36	19.21	2.04

Table 1: Results for the KITTI 2012 [10], the KITTI 2015 [10] and the MPI Sintel [6] training datasets. The listed error measures are the average endpoint error (AEE) and the percentage of erroneous pixels (BP) with a threshold of 3px.

KITTI 2012	Out-Noc	Out-All	Avg-Noc	Avg-All	KITTI 2015	Fl-bg	Fl-fg	Fl-all
ImpPB+SPCI	4.65 %	13.47 %	1.1 px	2.9 px	FlowNet2	10.75 %	8.75 %	10.41 %
FlowNet2	4.82 %	8.80 %	1.0 px	1.8 px	DCFlow	13.10 %	23.70 %	14.86 %
FlowFieldCNN	4.89 %	13.01 %	1.2 px	3.0 px	SOF	14.63 %	22.83 %	15.99 %
RicFlow [10]	4.96 %	13.04 %	1.3 px	3.2 px	<b>DF+OIR</b>	<b>15.11 %</b>	<b>23.45 %</b>	<b>16.50 %</b>
FlowFields+	5.06 %	13.14 %	1.2 px	3.0 px	ImpPB+SPCI	17.25 %	20.44 %	17.78 %
<b>DF+OIR</b>	<b>5.17 %</b>	<b>10.43 %</b>	<b>1.1 px</b>	<b>2.9 px</b>	FlowFieldCNN	18.33 %	20.42 %	18.68 %
PatchBatch [10]	5.29 %	14.17 %	1.3 px	3.3 px	RicFlow [10]	18.73 %	19.09 %	18.79 %
SODA-Flow [10]	5.57 %	10.71 %	1.3 px	2.8 px	FlowFields+	19.51 %	21.26 %	19.80 %
OAR-Flow [10]	5.69 %	10.72 %	1.4 px	2.8 px	PatchBatch [10]	19.98 %	26.50 %	21.07 %
DDF	5.73 %	14.18 %	1.4 px	3.4 px	DDF	20.36 %	25.19 %	21.17 %
PH-Flow	5.76 %	10.57 %	1.3 px	2.9 px	SODA-Flow [10]	20.01 %	29.14 %	21.53 %
FlowFields [10]	5.77 %	14.01 %	1.4 px	3.5 px	DiscreteFlow [10]	21.53 %	21.76 %	21.57 %
CPM-Flow [10]	5.79 %	13.70 %	1.3 px	3.2 px	OAR-Flow [10]	20.62 %	27.67 %	21.79 %
NLTGV-SC [10]	5.93 %	11.96 %	1.6 px	3.8 px	CPM-Flow [10]	22.32 %	22.81 %	22.40 %
DDS-DF	6.03 %	13.08 %	1.6 px	4.2 px	FullFlow [8]	23.09 %	24.79 %	23.37 %
TGV2ADCSIFT	6.20 %	15.15 %	1.5 px	4.5 px	SPM-BP	24.06 %	24.97 %	24.21 %
S2F-IF	6.20 %	15.68 %	1.4 px	3.5 px	EpicFlow [10]	25.81 %	28.69 %	26.29 %
DiscreteFlow [10]	6.23 %	16.63 %	1.3 px	3.6 px	DeepFlow [10]	27.96 %	31.06 %	28.48 %
BTF-ILLUM [10]	6.52 %	11.03 %	1.5 px	2.8 px	HS	39.90 %	51.39 %	41.81 %
EpicFlow [10]	7.88 %	17.08 %	1.5 px	3.8 px	DB-TV-L1	47.52 %	48.27 %	47.64 %

Table 2: Results for the KITTI 2012 [10] and KITTI 2015 [10] test datasets. Top non-anonymous pure optical flow methods, excluding methods that rely on additional information, such as stereo images, extra time-frames, semantic information or assume an underlying epipolar geometry, and related methods.

**Comparison to the Literature.** Finally, we evaluate our new order-adaptive variational refinement strategy on the withhold test datasets of the KITTI 2012 benchmark [10], the KITTI 2015 benchmark [10] and the MPI Sintel benchmark [6], by uploading the computed flow field to the online evaluation servers. We submitted the best performing setting, i.e. the combination of DiscreteFlow (DiscreteFlow matches + filtering + inpainting) with our order-adaptive refinement strategy. For a convenient overview we provide the results in Table 2 and Table 3, where we only listed non-anonymous pure optical flow methods that do not rely on additional information, such as stereo images, extra time-frames, semantic information or assume an underlying epipolar geometry. Moreover, we added results from EpicFlow [23] and OAR-Flow [10], if not already present in the list of the top results, since these methods rely on the standard pipeline and the order adaptive regularisation, respectively. As one can see, our variational refinement not only improves the results compared to the original DiscreteFlow approach (DiscreteFlow matches + filtering + inpainting + EpicFlow refinement) and the standard EpicFlow pipeline (DeepMatches + filtering + inpainting + EpicFlow refinement), it also outperforms recent pure variational methods with full coarse-to-fine schemes such as SODA-Flow (second order regularisation), OAR-Flow (adaptive order regularisation), and BTF-Illum (illumination-aware data term). Moreover, with Rank 6 (KITTI 2012), Rank 4 (KITTI 2015) and Ranks 3 and 10 (MPI Sintel) in the above Tables, the novel refinement approach offers a favourable performance in all benchmarks. This demonstrates that combining good initial matches with a sophisticated variational refinement allows to further improve the results by combining the advantages of both techniques.

**Runtimes.** For a colour image pair of size  $1242 \times 375$  (KITTI 2015 [10]) our C/C++ implementation of the variational refinement step running on a single core with 3.40 GHz (Intel Core i7-2600 CPU) requires about 35 s ( $\eta = 0.90$ ), 50 s ( $\eta = 0.95$ ) and 70 s ( $\eta = 1.00$ ). The previous pipeline steps sum up to 12 s (CPM), 80 s (Deepmatches) and 120 s (DiscreteFlow).

MPI Sintel clean	all	matched	unmatched	MPI Sintel final	all	matched	unmatched
FlowFields+	3.102	0.820	21.718	DCFlow	5.119	2.283	28.228
CPM2	3.253	0.980	21.812	FlowFieldsCNN	5.363	2.303	30.313
<b>DiscreteFlow+OIR</b>	<b>3.331</b>	<b>0.942</b>	<b>22.817</b>	S2F-IF	5.417	2.549	28.795
S2F-IF	3.500	0.988	23.986	RicFlow [□]	5.620	2.765	28.907
SPM-BPv2	3.515	1.020	23.865	FlowFields+	5.707	2.684	30.356
DCFlow	3.537	1.103	23.394	DeepDiscreteFlow	5.728	2.623	31.042
RicFlow [□]	3.550	1.264	22.220	FlowNet2-ft-sintel	5.739	2.752	30.108
CPM-Flow [▣]	3.557	1.189	22.889	FlowFields [▣]	5.810	2.621	31.799
DiscreteFlow [▣]	3.567	1.108	23.626	SPM-BPv2	5.812	2.754	30.743
FullFlow [▣]	3.601	1.296	22.424	<b>DiscreteFlow+OIR</b>	<b>5.862</b>	<b>2.864</b>	<b>30.303</b>
PatchBatch+Inter	3.624	1.324	22.397	FullFlow [▣]	5.895	2.838	30.793
FlowFields [▣]	3.748	1.056	25.700	CPM-Flow [▣]	5.960	2.990	30.177
FlowFieldsCNN	3.778	0.996	26.469	FlowNet2	6.016	2.977	30.807
DeepDiscreteFlow	3.863	1.296	24.820	GlobalPatchCollider	6.040	2.938	31.309
FlowNet2	3.959	1.468	24.294	DiscreteFlow [▣]	6.077	2.937	31.685
EpicFlow [▣]	4.115	1.360	26.595	EpicFlow [▣]	6.285	3.060	32.564
OAR-Flow [▣]	6.227	2.760	34.455	OAR-Flow [▣]	8.179	4.578	37.525

Table 3: Results for the MPI Sintel [6] test datasets in terms of the average endpoint error (AEE). Top non-anonymous optical flow methods and related methods.

**Limitations.** Our approach not only enables an accurate refinement of flow fields but also is capable of correcting errors. Naturally, this error correction capability is limited when it comes to erroneous flow vectors of small objects that undergo a large displacement. Such errors cannot be corrected if no correct matches are captured during the matching phase. Another limiting scenario can be observed at motion boundaries between foreground objects and homogeneous background regions. In this case it can appear that, despite the edge enhancing and edge preserving penaliser functions, the smoothness term over-smooths the edge. Additional segmentation information may help in this context.

## 7 Conclusion and Outlook

In this paper we proposed a novel variational refinement strategy for optical flow estimation. By combining an illumination-aware data term, that can keep up with many feature descriptors regarding their robustness, with an order-adaptive smoothness term, that locally selects between first and second order regularisation, we introduced recent concepts from the field of variational optical estimation into pipeline approaches that rely on variational refinement. In order to benefit from a good initialisation while still being able to correct errors, we also proposed a reduced coarse-to-fine scheme that start the computation at an intermediate level. Consistently good results on recent optical flow benchmarks showed that our novel variational refinement strategy not only allows to improve results compared to traditional refinement schemes, but also that it allows to outperform pure variational methods.

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

## References

- [1] C. Bailer, B. Taetz, and D. Stricker. Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation. In *Proc. IEEE International*

*Conference on Computer Vision*, pages 4015–4023, 2015.

- [2] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. Patchmatch: a randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics*, 28(3):24:1–24:11, 2009.
- [3] J. Braux-Zin, R. Dupont, and A. Bartoli. A general dense image matching framework combining direct and feature-based costs. In *Proc. IEEE International Conference on Computer Vision*, pages 185–192, 2013.
- [4] T. Brox and J. Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):500–513, 2011.
- [5] T. Brox, A. Bruhn, N. Papenbergh, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *Proc. European Conference on Computer Vision*, pages 25–36, 2004.
- [6] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *Proc. European Conference on Computer Vision*, pages 611–625, 2012.
- [7] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–311, 1997.
- [8] Q. Chen and V. Koltun. Full flow: Optical flow estimation by global optimization over regular grids. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 4706–4714, 2016.
- [9] O. Demetz, D. Hafner, and J. Weickert. The complete rank transform: A tool for accurate and morphologically invariant matching of structures. In *Proc. British Machine Vision Conference*, 2013.
- [10] O. Demetz, M. Stoll, S. Volz, J. Weickert, and A. Bruhn. Learning brightness transfer functions for the joint recovery of illumination changes and optical flow. In *Proc. European Conference on Computer Vision*, pages 455–471, 2014.
- [11] D. Gadot and L. Wolf. Patchbatch: a batch augmented loss for optical flow. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 4236–4245, 2016.
- [12] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the KITTI vision benchmark suite. In *In Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012.
- [13] M.D. Grossberg and S.K. Nayar. Modeling the space of camera response functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1272–1282, 2004.
- [14] D. Hafner, C. Schroers, and J. Weickert. Introducing maximal anisotropy into second order coupling models. In *In Proc. German Conference on Pattern Recognition*, pages 79–90, 2015.

- [15] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial intelligence*, 17 (1-3):185–203, 1981.
- [16] Y. Hu, R. Song, and Y. Li. Efficient coarse-to-fine patchmatch for large displacement optical flow. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 5704–5712, 2016.
- [17] Y. Hu, Y. Li, and R. Song. Robust interpolation of correspondences for large displacement optical flow. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (to appear)*, 2017.
- [18] C. Liu, J. Yuen, and A. Torralba. SIFT flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33 (5):978–994, 2011.
- [19] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [20] D. Maurer, M. Stoll, and A. Bruhn. Order-adaptive regularisation for variational optical flow: Global, local and in between. In *Proc. International Conference on Scale Space and Variational Methods in Computer Vision*, pages 550–562, 2017.
- [21] D. Maurer, M. Stoll, S. Volz, P. Gairing, and A. Bruhn. A comparison of isotropic and anisotropic second order regularisers for optical flow. In *Proc. International Conference on Scale Space and Variational Methods in Computer Vision*, pages 537–549, 2017.
- [22] M. Menze and A. Geiger. Object scene flow for autonomous vehicles. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 3061–3070, 2015.
- [23] M. Menze, C. Heipke, and A. Geiger. Discrete optimization for optical flow. In *Proc. German Conference on Pattern Recognition*, pages 16–28, 2015.
- [24] Y. Mileva, A. Bruhn, and J. Weickert. Illumination-robust variational optical flow with photometric invariants. In *Proc. German Conference on Pattern Recognition*, pages 152–162, 2007.
- [25] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2:629–639, 1990.
- [26] R. Ranftl, K. Bredies, and T. Pock. Non-local total generalized variation for optical flow estimation. In *Proc. European Conference on Computer Vision*, pages 439–454, 2014.
- [27] H. Rashwan, M. Mohamed, M. Garcia, B. Mertsching, and D. Puig. Illumination robust optical flow model based on histogram of oriented gradients. In *Proc. German Conference on Pattern Recognition*, pages 354–363, 2013.
- [28] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid. Epicflow: Edge-preserving interpolation of correspondences for optical flow. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1164–1172, 2015.

- [29] M. Stoll, S. Volz, and A. Bruhn. Adaptive integration of feature matches into variational optical flow methods. In *Proc. Asian Conference on Pattern Recognition*, pages 1–14, 2013.
- [30] M. Stoll, S. Volz, D. Maurer, and A. Bruhn. A time-efficient optimisation framework for parameters of optical flow methods. In *Proc. Scandinavian Conference on Image Analysis*, pages 41–53, 2017.
- [31] E. Tola, V. Lepetit, and P. Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5):815–830, 2010.
- [32] Z. Tu, R. Poppe, and R. C. Veltkamp. Weighted local intensity fusion method for variational optical flow estimation. *Pattern Recognition*, 50:223–232, 2016.
- [33] A. Wedel, D. Cremers, T. Pock, and H. Bischof. Structure-and motion-adaptive regularization for high accuracy optic flow. In *Proc. IEEE International Conference on Computer Vision*, pages 1663–1668, 2009.
- [34] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid. Deepflow: Large displacement optical flow with deep matching. In *Proc. IEEE International Conference on Computer Vision*, pages 1385–1392, 2013.
- [35] M. Werlberger, T. Pock, and H. Bischof. Motion estimation with non-local total variation regularization. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 2464–2471, 2010.
- [36] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34:1744–1757, 2012.
- [37] H. Zimmer, A. Bruhn, and J. Weickert. Optic flow in harmony. *International Journal of Computer Vision*, 93(3):368–388, 2011.