Universität des Saarlandes

Fachrichtung 6.1 – Mathematik

Preprint Nr. 272

# Optic Flow in Harmony

Henning Zimmer, Andrés Bruhn and Joachim Weickert

Saarbrücken 2010

Fachrichtung 6.1 – Mathematik
Universität des Saarlandes

Preprint No. 272
submitted: November 25, 2010

# Optic Flow in Harmony

**Henning Zimmer**
Mathematical Image Analysis Group
Dept. of Mathematics and Computer Science
Saarland University, Campus E1.1, 66041 Saarbrücken, Germany
zimmer@mia.uni-saarland.de

**Andrés Bruhn**
Mathematical Image Analysis Group
Dept. of Mathematics and Computer Science
Saarland University, Campus E1.1, 66041 Saarbrücken, Germany
bruhn@mia.uni-saarland.de

**Joachim Weickert**
Mathematical Image Analysis Group
Dept. of Mathematics and Computer Science
Saarland University, Campus E1.1, 66041 Saarbrücken, Germany
weickert@mia.uni-saarland.de

Edited by
FR 6.1 – Mathematik
Universität des Saarlandes
Postfach 15 11 50
66041 Saarbrücken
Germany

Fax: + 49 681 302 4443
e-Mail: preprint@math.uni-sb.de
WWW: http://www.math.uni-sb.de/

# Optic Flow in Harmony

Henning Zimmer Andrés Bruhn Joachim Weickert

**Abstract**

Most variational optic flow approaches just consist of three constituents: a data term, a smoothness term and a smoothness weight. In this paper, we present an approach that harmonises these three components. We start by developing an advanced data term that is robust under outliers and varying illumination conditions. This is achieved by using constraint normalisation, and an HSV colour representation with higher order constancy assumptions and a separate robust penalisation. Our novel anisotropic smoothness is designed to work complementary to the data term. To this end, it incorporates directional information from the data constraints to enable a filling-in of information solely in the direction where the data term gives no information, yielding an optimal complementary smoothing behaviour. This strategy is applied in the spatial as well as in the spatio-temporal domain. Finally, we propose a simple method for automatically determining the optimal smoothness weight. This method bases on a novel concept that we call "optimal prediction principle" (OPP). It states that the flow field obtained with the optimal smoothness weight allows for the best prediction of the next frames in the image sequence. The benefits of our "optic flow in harmony" (OFH) approach are demonstrated by an extensive experimental validation and by a competitive performance at the widely used Middlebury optic flow benchmark.

## 1 Introduction

Despite almost three decades of research on variational optic flow approaches, there have been hardly any investigations on the compatibility of their three main components: the data term, the smoothness term and the smoothness weight. While the data term models constancy assumptions on image features, the smoothness term penalises fluctuations in the flow field, and the smoothness weight determines the balance between the two terms. In this paper, we present the *optic flow in harmony (OFH)* method, which harmonises the three constituents by following two main ideas:
*(i)* Widely-used data terms such as the one resulting from the linearised brightness constancy assumption only constrain the flow in one direction.

However, most smoothness terms impose smoothness also in the data constraint direction, leading to an undesirable interference. A notable exception is the anisotropic smoothness term proposed by Nagel and Enkelmann [27]. At large image gradients, their regulariser solely smoothes along image edges. For a basic data term modelling the brightness constancy assumption, this smoothing direction is orthogonal to the data constraint direction and thus both term complement each other in an optimal manner. Unfortunately, this promising concept of complementarity between data and smoothness term has not been further investigated after 1986. Our paper revives this concept for state-of-the-art optic flow models by presenting a novel *complementary smoothness term* in conjunction with an advanced data term.
*(ii)* Having adjusted the smoothing behaviour to the imposed data constraints, it remains to determine the optimal balance between the two terms for the image sequence under consideration. This comes down to selecting an appropriate smoothness weight, which is usually considered a difficult task. We propose a method that is easy to implement for all variational optic flow approaches and nevertheless gives surprisingly good results. It bases on the assumption that the flow estimate obtained by an optimal smoothness weight allows for the best possible prediction of the next frames in the image sequence. This novel concept we name *optimal prediction principle (OPP)*.

## 1.1 Related Work

In the first ten years of research on optic flow, several basic strategies have been considered, e.g. phase-based methods [15], local methods [23, 5] and energy-based methods [19]. In recent years, the latter class of methods became increasingly popular, mainly due to their potential for giving highly accurate results. Within energy-based methods, one can distinguish discrete approaches that minimise a discrete energy function and are often probabilistically motivated, and variational approaches that minimise a continuous energy functional.
Our variational approach naturally incorporates concepts that have proven their benefits over the years. In the following, we briefly review advances in the design of data and smoothness terms that are influential for our work.

**Data Terms.** To cope with outliers caused by noise or occlusions, Black and Anandan [6] replaced the quadratic penalisation from Horn and Schunck [19] by a robust subquadratic penaliser.
To obtain robustness under additive illumination changes, Brox et al. [8] combined the classical brightness constancy assumption [19] with the higher-order gradient constancy assumption [37, 31]. Bruhn and Weickert [9] im-

proved this idea by a separate robust penalisation of brightness and gradient constancy assumption. This gives advantages if one of the two constraints produces an outlier. Recently, Xu et al. [45] went a step further and estimate a binary map that locally selects between imposing either brightness or gradient constancy. An alternative to higher-order constancy assumptions can be to preprocess the images by a structure-texture decomposition [40].
In addition to additive illumination changes, realistic scenarios also encompass a multiplicative part [38]. For colour image sequences, this issue can be tackled by normalising the colour channels [17], or by using alternative colour spaces with photometric invariances [17, 38, 24]. If one is restricted to greyscale sequences, using log-derivatives [24] can be useful.
A further successful modification of the data term has been reported by performing a constraint normalisation [34, 21, 32]. It prevents an undesirable overweighting of the data term at large image gradient locations.

**Smoothness Terms.** First ideas go back to Horn and Schunck [19] who used a *homogeneous* regulariser that does not respect any flow discontinuities. Since different image objects may move in different directions, it is, however, desirable to also permit discontinuities.
This can for example be achieved by using *image-driven* regularisers that take into account image discontinuities. Alvarez et al. [1] proposed an isotropic model with a scalar-valued weight function that reduces the regularisation at image edges. An anisotropic counterpart that also exploits the directional information of image discontinuities was introduced by Nagel and Enkelmann [27]. Their method regularises the flow field along image edges but not across them. Note that for a basic data term modelling the brightness constancy assumption, the image edge direction coincides with the complementary direction orthogonal to the data constraint direction.
Of course, not every image edge will coincide with a flow edge. Thus, image-driven strategies are prone to give oversegmentation artefacts in textured image regions. To avoid this, *flow-driven* regularisers have been proposed that respect discontinuities of the evolving flow field and are therefore not misled by image textures. In the isotropic setting this comes down to the use of robust, subquadratic penalisers which are closely related to line processes [7]. For energy-based optic flow methods, such a strategy was used in [33, 31]. An anisotropic extension was later presented by Weickert and Schnörr [42]. The drawback of flow-driven regularisers lies in less well-localised flow edges compared to image-driven approaches.
Concerning the individual problems of image- and flow-driven strategies, the idea arises to combine the advantages of both worlds. This goal was first

achieved in the discrete method of Sun et al. [36]. There, the authors developed an anisotropic regulariser that uses directional flow derivatives steered by image structures. This allows to adapt the smoothing direction to the direction of image structures and the smoothing strength to the flow contrast. We call such a strategy *image- and flow-driven* regularisation. It combines the benefits of image- and flow-driven methods, i.e. sharp flow edges without oversegmentation problems.
The smoothness terms discussed so far only assume smoothness of the flow field in the spatial domain. As image sequences usually consist of more than two frames, yielding more than one flow field, it makes sense to also assume temporal smoothness of the flow fields. This leads to *spatio-temporal* smoothness terms. In a discrete setting they go back to Murray and Buxton [25]. For variational approaches, an image-driven spatio-temporal smoothness terms was proposed by Nagel [26] and a flow-driven counterpart was later presented by Weickert and Schnörr [43].

**Automatic Parameter Selection.** It is well-known that an appropriate choice of the smoothness weight is crucial for obtaining favourable results. Nevertheless, there has been remarkably little research on methods that automatically estimate the optimal smoothness weight or other model parameters.
Concerning an optimal selection of the smoothness weight for variational optic flow approaches, Ng and Solo [28] proposed an error measure which can be estimated from the image sequence and the flow estimate only. Using this measure, a brute-force search for the smoothness weight that gives the smallest error is performed. Computing the proposed error measure is, however, computationally expensive, especially for robust data terms. Ng and Solo [28] hence restricted their focus to the basic method of Horn and Schunck [19]. In a Bayesian framework, a parameter selection approach that can also handle robust data terms was presented by Krajsek and Mester [20]. This method jointly estimates the flow and the model parameters where the latter encompass the smoothness weight and also the relative weights of different data terms. This method does not require a brute-force search, but the minimisation of the objective function is more complicated and only computationally feasible if certain approximations are performed.

## 1.2 Our Contributions

The OFH method is obtained in three steps. We first develop a robust and invariant data term. Then an anisotropic image- and flow-driven smoothness

term is designed that works complementary to the data term. Finally we propose a simple method for automatically determining the optimal smoothness weight for the given image sequence.
Our data term combines the brightness and the gradient constancy assumption, and performs a constraint normalisation. It further uses a Hue-Saturation-Value (HSV) colour representation with a separate robustification of each channel. The latter is motivated by the fact that each channel in the HSV space has a distinct level of photometric invariance and information content. Hence, a separate robustification allows to choose the most reliable channel at each position. Our anisotropic complementary smoothness term takes into account directional information from the constraints imposed in the data term. Across *"constraint edges"*, we perform a robust penalisation to reduce the smoothing in the direction where the data term gives the most information. Along constraint edges, where the data term gives no information, a strong filling-in by using a quadratic penalisation makes sense. This strategy not only allows for an optimal complementarity between data and smoothness term, but also leads to a desirable image- and flow-driven behaviour. We further show that our regulariser can easily be extended to work in the spatio-temporal domain. Our method for determining the optimal smoothness weight bases on the proposed OPP concept. This results in finding the optimal smoothness weight as the one corresponding to the flow field with the best prediction quality. To judge the latter, we evaluate the data constraints between the first and the third frame of the sequence. Under mild assumptions (constant speed, linear trajectory of objects) this can be realised by simply doubling the flow vectors. Due to its simplicity, our method is easy to implement for all variational optic flow approaches, but nevertheless produces surprisingly good results.
The present article extends our shorter conference paper [46] by the following points: *(i)* A more extensive derivation and discussion of the data term. *(ii)* An explicit discussion on the adequate treatment of the hue channel of the HSV colour space. *(iii)* A taxonomy of existing smoothness terms within a novel general framework. The latter allows to reformulate most existing as well as our novel regulariser in a common notation. *(iv)* The extension of our complementary regulariser to the spatio-temporal domain. *(v)* A simple method for automatically selecting the smoothness weight. *(vi)* A deeper discussion of implementation issues. *(vii)* A more extensive experimental validation.

**Organisation.** In Section 2 we present our variational optic flow model with the robust data term and the complementary smoothness term. The

latter is then extended to the spatio-temporal domain. Section 3 describes the method proposed for determining the smoothness weight. After discussing implementation issues in Section 4, we show experiments in Section 5. The paper is finished with conclusions and an outlook on future work in Section 6.

# 2 Variational Optic Flow

Let $f(\mathbf{x})$ be a grayscale image sequence where $\mathbf{x} := (x, y, t)^\top$. Here, the vector $(x, y)^\top \in \Omega$ denotes the location within a rectangular image domain $\Omega \subset \mathbb{R}^2$, and $t \in [0, T]$ denotes time. We further assume that $f$ is presmoothed by a Gaussian convolution: Given an image sequence $f_0(\mathbf{x})$, we obtain $f(\mathbf{x}) = (K_\sigma * f_0)(\mathbf{x})$, where $K_\sigma$ is a spatial Gaussian of standard deviation $\sigma$ and $*$ denotes the convolution operator.
The optic flow field $\mathbf{w} := (u, v, 1)^\top$ describes the displacement vector field between two frames at time $t$ and $t + 1$. It is found by minimising a global energy functional of the general form

$$E(u, v) = \int_\Omega \big(M(u, v) + \alpha\, V(\nabla_2 u, \nabla_2 v)\big)\, \mathrm{d}x\, \mathrm{d}y \ , \tag{1}$$

where $\nabla_2 := (\partial_x, \partial_y)^\top$ denotes the spatial gradient operator. The term $M(u, v)$ denotes the data term, $V(\nabla_2 u, \nabla_2 v)$ the smoothness term, and $\alpha > 0$ is a smoothness weight. Note that the energy (1) refers to the spatial case where one computes one flow field between two frames at time $t$ and $t + 1$. The more general spatio-temporal case that uses all frames $t \in [0, T]$ will be presented in Section 2.3.
According to the calculus of variations [14], a minimiser $(u, v)$ of the energy (1) necessarily has to fulfil the associated Euler-Lagrange equations

$$\begin{aligned} \partial_u M - \alpha \left(\partial_x \left(\partial_{u_x} V\right) + \partial_y \left(\partial_{u_y} V\right)\right) &= 0 \ , & (2)\\ \partial_v M - \alpha \left(\partial_x \left(\partial_{v_x} V\right) + \partial_y \left(\partial_{v_y} V\right)\right) &= 0 & (3) \end{aligned}$$

with homogeneous Neumann boundary conditions.

## 2.1 Data Term

Let us now derive our data term in a systematic way. The classical starting point is the *brightness constancy assumption* used by Horn and Schunck [19]. It states that image intensities remain constant under their displacement, i.e. $f(\mathbf{x}+\mathbf{w}) = f(\mathbf{x})$. Assuming that the image sequence is smooth and that the

displacements are small, we can perform a first-order Taylor expansion that yields the linearised *optic flow constraint (OFC)*

$$0 = f_x\, u + f_y\, v + f_t = \nabla_3^\top f\ \mathbf{w}\ \ , \tag{4}$$

where $\nabla_3 := (\partial_x, \partial_y, \partial_t)^\top$ is the spatio-temporal gradient operator and subscripts denote partial derivatives. With a quadratic penalisation, the corresponding data term is given by

$$M_1(u,v) = \left(\nabla_3^\top f\ \mathbf{w}\right)^2 = \mathbf{w}^\top \mathbf{J}_0\, \mathbf{w}\ \ , \tag{5}$$

with the tensor

$$\mathbf{J}_0 := \nabla_3 f\ \nabla_3^\top f\ \ . \tag{6}$$

The single equation given by the OFC involves two unknowns $u$ and $v$. It is thus not sufficient to compute a unique solution, which is known as the aperture problem [4]. Nevertheless, the OFC does allow to compute the flow component orthogonal to image edges, the so-called *normal flow*. For $|\nabla_2 f| \neq 0$ it is defined as

$$\mathbf{w}_n := \left(\mathbf{u}_n^\top, 1\right)^\top := \left(-\frac{f_t}{|\nabla_2 f|}\frac{\nabla_2^\top f}{|\nabla_2 f|},\ 1\right)^\top\ \ . \tag{7}$$

**Normalisation.** Our experiments will show that normalising the data term can be beneficial. Following [34, 21, 32] and using the abbreviation $\mathbf{u} := (u,v)^\top$, we rewrite the data term $M_1$ as

$$\begin{aligned} M_1(u,v) &= \left(\nabla_2^\top f\ \mathbf{u} + f_t\right)^2 = \left[|\nabla_2 f|\left(\frac{\nabla_2^\top f\ \mathbf{u}}{|\nabla_2 f|} + \frac{f_t}{|\nabla_2 f|}\right)\right]^2 \\ &= |\nabla_2 f|^2\left[\frac{\nabla_2^\top f}{|\nabla_2 f|}\left(\mathbf{u} + \frac{f_t\,\nabla_2 f}{|\nabla_2 f|^2}\right)\right]^2 \\ &= |\nabla_2 f|^2\Bigg(\underbrace{\frac{\nabla_2^\top f}{|\nabla_2 f|}\left(\mathbf{u} - \mathbf{u}_n\right)}_{=:d}\Bigg)^2\ \ . \end{aligned} \tag{8}$$

The term $d$ constitutes a projection of the difference between the estimated flow $\mathbf{u}$ and the normal flow $\mathbf{u}_n$ in the direction of the image gradient $\nabla_2 f$. In a geometric interpretation, the term $d$ describes the distance from $\mathbf{u}$ to the line $l$ in the $uv$-space that is given by

$$v = -\frac{f_x}{f_y}\, u - \frac{f_t}{f_y}\ \ . \tag{9}$$

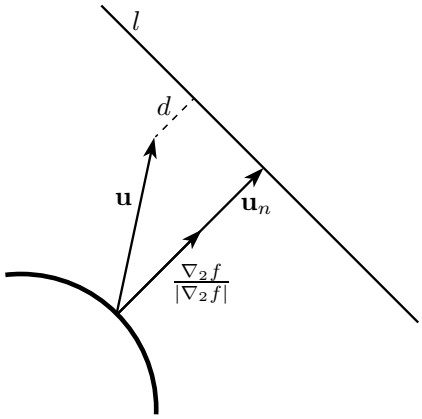


Figure 1: Geometric interpretation of the rewritten data term (8).

On this line, the flow $\mathbf{u}$ has to lie according to the OFC (4), and the normal flow $\mathbf{u}_n$ is the smallest vector that lies on this line. A sketch of this situation is given in Figure 1. Our geometric interpretation suggests that one should ideally penalise the distance $d$ in a data term $M_2(u, v) = d^2$. The data term $M_1$, however, weighs this distance by the squared spatial image gradient, as $M_1(u, v) = |\nabla_2 f|^2 \, d^2$, see (8). This results in a stronger enforcement of the data constraint at high gradient locations. This overweighting is undesirable as large gradients can be caused by unreliable structures, such as noise or occlusions.

As a remedy, we normalise the data term $M_1$ by multiplying it with a factor [34, 21]

$$\theta_0 := \frac{1}{|\nabla_2 f|^2 + \zeta^2} \ , \tag{10}$$

where the regularisation parameter $\zeta > 0$ avoids division by zero and additionally reduces the influence of small gradients, while not influencing the normalisation for gradients significantly larger than $\zeta^2$. A normalised version of $M_1$ can then be written as

$$M_2(u, v) = \mathbf{w}^\top \bar{\mathbf{J}}_0 \, \mathbf{w} \ , \tag{11}$$

with the normalised tensor

$$\bar{\mathbf{J}}_0 := \theta_0 \, \mathbf{J}_0 = \theta_0 \left( \nabla_3 f \, \nabla_3^\top f \right) \ . \tag{12}$$

**Gradient Constancy Assumption.** To render the data term robust under additive illumination changes, it was proposed to impose the *gradient constancy assumption* [37, 31, 8]. In contrast to the brightness constancy

assumption, it states that image *gradients* remain constant under their displacement, i.e. $\nabla_2 f(\mathbf{x}+\mathbf{w}) = \nabla_2 f(\mathbf{x})$. A Taylor linearisation gives

$$\nabla_3^\top f_x \, \mathbf{w} = 0 \; , \quad \text{and} \quad \nabla_3^\top f_y \, \mathbf{w} = 0 \; , \tag{13}$$

respectively. Combining both brightness and gradient constancy assumption in a quadratic way gives the data term

$$M_3(u,v) = \mathbf{w}^\top \mathbf{J} \, \mathbf{w} \; , \tag{14}$$

where the tensor $\mathbf{J}$ can be written in the *motion tensor notation* [10] that allows to combine the two constancy assumptions in a joint tensor

$$\begin{aligned} \mathbf{J} &:= \mathbf{J}_0 + \gamma \, \mathbf{J}_{xy} := \mathbf{J}_0 + \gamma \, (\mathbf{J}_x + \mathbf{J}_y) \\ &:= \nabla_3 f \, \nabla_3^\top f + \gamma \left( \nabla_3 f_x \, \nabla_3^\top f_x + \nabla_3 f_y \, \nabla_3^\top f_y \right) \; , \end{aligned} \tag{15}$$

where the parameter $\gamma > 0$ steers the contribution of the gradient constancy assumption.
To normalise $M_3$, we replace the motion tensor $\mathbf{J}$ by its normalised counterpart

$$\begin{aligned} \bar{\mathbf{J}} &:= \bar{\mathbf{J}}_0 + \gamma \, \bar{\mathbf{J}}_{xy} := \bar{\mathbf{J}}_0 + \gamma \left( \bar{\mathbf{J}}_x + \bar{\mathbf{J}}_y \right) \\ &:= \theta_0 \, \mathbf{J}_0 + \gamma \, (\theta_x \, \mathbf{J}_x + \theta_y \, \mathbf{J}_y) \\ &:= \quad \theta_0 \big( \nabla_3 f \; \nabla_3^\top f \big) \\ &\quad + \gamma \left( \theta_x \big( \nabla_3 f_x \; \nabla_3^\top f_x \big) + \theta_y \big( \nabla_3 f_y \; \nabla_3^\top f_y \big) \right) \; , \end{aligned} \tag{16}$$

with two additional normalisation factors defined as

$$\theta_x := \frac{1}{|\nabla_2 f_x|^2 + \zeta^2} \; , \quad \text{and} \quad \theta_y := \frac{1}{|\nabla_2 f_y|^2 + \zeta^2} \; . \tag{17}$$

The normalised data term $M_4$ is given by

$$M_4(u,v) = \mathbf{w}^\top \bar{\mathbf{J}} \, \mathbf{w} \; . \tag{18}$$

**Colour Image Sequences.** In a next step we extend our data term to multi-channel sequences $(f^1(\mathbf{x}), f^2(\mathbf{x}), f^3(\mathbf{x}))$. If one uses the standard RGB colour space, the three channels represent the red, green and blue channel,

respectively. We couple the three colour channels in the motion tensor

$$\begin{aligned}\bar{\mathbf{J}}^{\,c} &:= \sum_{i=1}^{3} \bar{\mathbf{J}}^{\,i} := \sum_{i=1}^{3} \left[\bar{\mathbf{J}}_0^{\,i} + \gamma\, \bar{\mathbf{J}}_{xy}^{\,i}\right] := \sum_{i=1}^{3} \left[\bar{\mathbf{J}}_0^{\,i} + \gamma \left(\bar{\mathbf{J}}_x^{\,i} + \bar{\mathbf{J}}_y^{\,i}\right)\right] \\ &:= \sum_{i=1}^{3} \Big[\theta_0^i \left(\nabla_3 f^i\ \nabla_3^\top f^i\right) \\ &\quad + \gamma\ \left(\theta_x^i \left(\nabla_3 f_x^i\ \nabla_3^\top f_x^i\right) + \theta_y^i \left(\nabla_3 f_y^i\ \nabla_3^\top f_y^i\right)\right)\Big]\ ,\end{aligned} \tag{19}$$

with normalisation factors $\theta^i$ for each colour channel $f^i$. The corresponding data term reads as

$$M_5(u,v) = \mathbf{w}^\top \bar{\mathbf{J}}^{\,c}\, \mathbf{w}\ . \tag{20}$$

**Photometric Invariant Colour Spaces.** Realistic illumination models encompass a multiplicative influence [38], which cannot be captured by the gradient constancy assumption that is only invariant under additive illumination changes. This problem can be tackled by using the *Hue Saturation Value (HSV)* colour space, as proposed in [17]. The hue channel is invariant under global and local multiplicative illumination changes, as well as under local additive changes. The saturation channel is only invariant under global multiplicative illumination changes, and the value channel exhibits no invariances. Mileva et al. [24] thus only used the hue channel for optic flow computation as it exhibits the most invariances. We will additionally use the saturation and value channel, because they contain information that is not encoded in the hue channel.

As an example, consider the HSV decomposition of the *Rubberwhale* image shown in Figure 2. As we can see, the shadow at the left of the wheel (shown in the zoom) is not present in the hue and the saturation channel, but appears in the value channel. Nevertheless, especially the hue channel discards a lot of image information, as can be observed for the striped cloth. This information is, on the other hand, available in the value channel.

One problem when using a HSV colour representation is that the hue channel $f^1$ describes an angle in a colour circle, i.e. $f^1 \in [0°, 360°)$. The hue channel is hence not differentiable at the interface between 0° and 360°. Our solution to this problem is to consider the unit vector $(\cos f^1, \sin f^1)^\top$ corresponding to the angle $f^1$. This results in treating the hue channel as two (coupled) channels, which are both differentiable. The corresponding motion tensor for the brightness constancy assumption consequently reads as

$$\bar{\mathbf{J}}_0^{\,1} := \theta_0^1 \left(\nabla_3 \cos f^1\ \nabla_3^\top \cos f^1 + \nabla_3 \sin f^1\ \nabla_3^\top \sin f^1\right)\ , \tag{21}$$

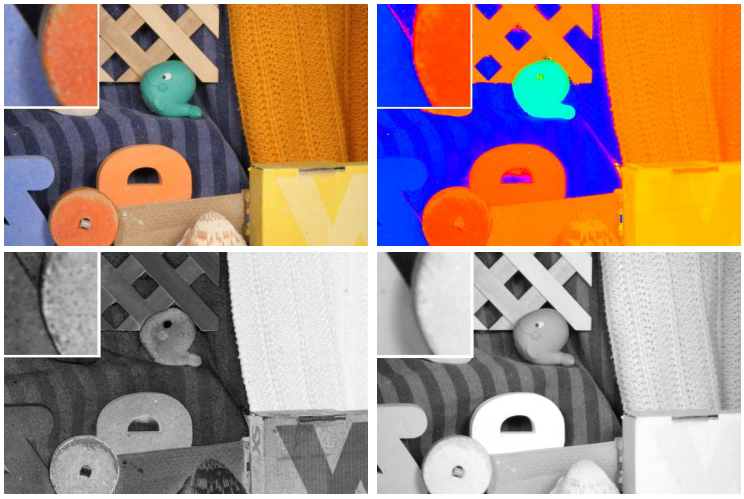

Figure 2: HSV decomposition on the example of the *Rubberwhale* image from the Middlebury optic flow database [2] (*http://vision.middlebury.edu/flow/data/*). *First row, from left to right:* **(a)** Colour image with zoom in shadow region left of the wheel. **(b)** Hue channel, visualised with full saturation and value. *Second row, from left to right:* **(c)** Saturation channel. **(d)** Value channel

where the normalisation factor is defined as

$$\theta_0^1 := \frac{1}{|\nabla_2 \cos f^1|^2 + |\nabla_2 \sin f^1|^2 + \zeta^2} \quad . \tag{22}$$

The tensor $\bar{\mathbf{J}}_{xy}^{\,1}$ for the gradient constancy assumption is adapted accordingly. Note that in the *differentiable parts* of the hue channel, the motion tensor (21) is equivalent to our earlier definition, as

$$\begin{aligned} & \nabla_3 \cos f^1 \; \nabla_3^\top \cos f^1 \; + \; \nabla_3 \sin f^1 \; \nabla_3^\top \sin f^1 \\ = \; & \sin^2 f^1 \left( \widetilde{\nabla}_3 f^1 \; \widetilde{\nabla}_3^\top f^1 \right) + \cos^2 f^1 \left( \widetilde{\nabla}_3 f^1 \; \widetilde{\nabla}_3^\top f^1 \right) \\ = \; & \widetilde{\nabla}_3 f^1 \widetilde{\nabla}_3^\top f^1 \quad , \end{aligned} \tag{23}$$

where $\widetilde{\nabla}$ denotes the gradient in the differentiable parts of the hue channel.

**Robust Penalisers.** To provide robustness of the data term against outliers caused by noise and occlusions, Black and Anandan [6] proposed to refrain from a quadratic penalisation. Instead they use a subquadratic penalisation function $\Psi_M(s^2)$, where $s^2$ denotes the quadratic data term. Using

such a robust penaliser within our data term yields

$$M_6(u,v) = \Psi_M\big(\mathbf{w}^\top \bar{\mathbf{J}}^{\,c}\, \mathbf{w}\big) \quad . \tag{24}$$

Good results are reported in [8] for the subquadratic penaliser $\Psi_M(s^2) := \sqrt{s^2+\varepsilon^2}$ using a small regularisation parameter $\varepsilon > 0$.
Bruhn and Weickert [9] later proposed a *separate* penalisation of the brightness and the gradient constancy assumption, which is advantageous if one assumption produces an outlier. Incorporating this strategy into our approach gives the data term

$$M_7(u,v) = \Psi_M\big(\mathbf{w}^\top \bar{\mathbf{J}}_0^{\,c}\, \mathbf{w}\big) + \gamma\, \Psi_M\big(\mathbf{w}^\top \bar{\mathbf{J}}_{xy}^{\,c}\, \mathbf{w}\big) \quad , \tag{25}$$

where the separate motion tensors are defined as

$$\bar{\mathbf{J}}_0^{\,c} := \sum_{i=1}^{3} \bar{\mathbf{J}}_0^{\,i} \, , \quad \text{and} \quad \bar{\mathbf{J}}_{xy}^{\,c} := \sum_{i=1}^{3} \bar{\mathbf{J}}_{xy}^{\,i} \quad . \tag{26}$$

We will go further by proposing a separate robustification of each colour channel in the HSV space. This can be justified by the distinct information content of each of the three channels, see Figure 2, that drives the optic flow estimation in different ways. The separate robustification then downweights the influence of less appropriate colour channels.

**Final Data Term.** Incorporating our separate robustification idea into $M_7$ brings us to our final data term

$$M(u,v) = \sum_{i=1}^{3} \Psi_M\big(\mathbf{w}^\top \bar{\mathbf{J}}_0^{\,i}\, \mathbf{w}\big) + \gamma \left( \sum_{i=1}^{3} \Psi_M\big(\mathbf{w}^\top \bar{\mathbf{J}}_{xy}^{\,i}\, \mathbf{w}\big) \right), \tag{27}$$

with the motion tensors $\bar{\mathbf{J}}^{\,1}$ adapted to the HSV colour space as described before. Note that our final data term is *(i)* normalised, *(ii)* combines the brightness and gradient constancy assumption, and *(iii)* uses the HSV colour space with *(iv)* a separate robustification of all colour channels.
The contributions of our data term (27) to the Euler-Lagrange equations (2)

and (3) are given by

$$\partial_u M = \sum_{i=1}^{3}\left(\Psi'_M\big(\mathbf{w}^\top \bar{\mathbf{J}}_0^i\,\mathbf{w}\big)\cdot\Big(\big[\bar{\mathbf{J}}_0^i\big]_{1,1}u+\big[\bar{\mathbf{J}}_0^i\big]_{1,2}v+\big[\bar{\mathbf{J}}_0^i\big]_{1,3}\Big)\right) \tag{28}$$
$$+\gamma\left(\sum_{i=1}^{3}\left(\Psi'_M\big(\mathbf{w}^\top \bar{\mathbf{J}}_{xy}^i\,\mathbf{w}\big)\cdot\Big(\big[\bar{\mathbf{J}}_{xy}^i\big]_{1,1}u+\big[\bar{\mathbf{J}}_{xy}^i\big]_{1,2}v+\big[\bar{\mathbf{J}}_{xy}^i\big]_{1,3}\Big)\right)\right),$$

$$\partial_v M = \sum_{i=1}^{3}\left(\Psi'_M\big(\mathbf{w}^\top \bar{\mathbf{J}}_0^i\,\mathbf{w}\big)\cdot\Big(\big[\bar{\mathbf{J}}_0^i\big]_{1,2}u+\big[\bar{\mathbf{J}}_0^i\big]_{2,2}v+\big[\bar{\mathbf{J}}_0^i\big]_{2,3}\Big)\right) \tag{29}$$
$$+\gamma\left(\sum_{i=1}^{3}\left(\Psi'_M\big(\mathbf{w}^\top \bar{\mathbf{J}}_{xy}^i\,\mathbf{w}\big)\cdot\Big(\big[\bar{\mathbf{J}}_{xy}^i\big]_{1,2}u+\big[\bar{\mathbf{J}}_{xy}^i\big]_{2,2}v+\big[\bar{\mathbf{J}}_{xy}^i\big]_{2,3}\Big)\right)\right),$$

where $[\mathbf{J}]_{m,n}$ denotes the entry in row $m$ and column $n$ of the tensor $\mathbf{J}$, and $\Psi'_M(s^2)$ denotes the derivative of $\Psi_M(s^2)$ w.r.t. its argument. Analysing the terms (28) and (29), we see that the separate robustification of the HSV channels makes sense: If a specific channel violates the imposed constancy assumption at a certain location, the corresponding argument of the decreasing function $\Psi'_M$ will be large, yielding a downweighting of this channel. Other channels that satisfy the constancy assumption then have a dominating influence on the solution. This will be confirmed by a specific experiment in Section 5.1.

## 2.2 Smoothness Term

Following the extensive taxonomy on optic flow regularisers [42], we sketch some existing smoothness terms that led to our novel complementary regulariser. We rewrite the regularisers in a novel framework that unifies their notation and eases their comparison.

**Preliminaries for the General Framework.** We first introduce concepts that will be used in our general framework.
Anisotropic image-driven regularisers take into account directional information from image structures. These information can be obtained by considering the structure tensor [16]

$$\mathbf{S}_\rho := K_\rho * \Big[\nabla_2 f\ \nabla_2^\top f\Big] =: \sum_{i=1}^{2}\mu_i\ \mathbf{s}_i\,\mathbf{s}_i^\top\ , \tag{30}$$

with an integration scale $\rho > 0$. The structure tensor is a symmetric, positive semidefinite $2\times 2$ matrix that possesses two orthonormal eigenvectors $\mathbf{s}_1$ and

$\mathbf{s}_2$ with corresponding eigenvalues $\mu_1 \geq \mu_2 \geq 0$. The vector $\mathbf{s}_1$ points across image structures, whereas the vector $\mathbf{s}_2$ points along them. In the case for $\rho = 0$, i.e. without considering neighbourhood information, one obtains

$$\mathbf{s}_1^0 = \frac{\nabla_2 f}{|\nabla_2 f|} \ , \quad \text{and} \quad \mathbf{s}_2^0 = \frac{\nabla_2^\perp f}{|\nabla_2 f|} \ , \tag{31}$$

where $\nabla_2^\perp f := (-f_y, f_x)^\top$ denotes the vector orthogonal to $\nabla_2 f$. For $\rho = 0$ we also have that $|\nabla_2 f|^2 = \text{tr } \mathbf{S}_0$, with tr denoting the trace operator.
Most regularisers impose smoothness by penalising the magnitude of the flow gradients. As $\mathbf{s}_1$ and $\mathbf{s}_2$ constitute an orthonormal basis, we can write

$$|\nabla_2 u|^2 \ = \ u_x^2 + u_y^2 \ = \ u_{\mathbf{s}_1}^2 + u_{\mathbf{s}_2}^2 \ , \tag{32}$$

using the *directional derivatives* $u_{\mathbf{s}_i} := \mathbf{s}_i^\top \nabla_2 u$. A corresponding rewriting can also be performed for $|\nabla_2 v|^2$.
To analyse the smoothing behaviour of the regularisers, we will consider the corresponding Euler-Lagrange equations that can be written in the form

$$\partial_u M - \alpha \ \text{div} \left( \mathbf{D} \, \nabla_2 u \right) \ = \ 0 \ , \tag{33}$$

$$\partial_v M - \alpha \ \text{div} \left( \mathbf{D} \, \nabla_2 v \right) \ = \ 0 \ , \tag{34}$$

with a diffusion tensor $\mathbf{D}$ that steers the smoothing of the flow components $u$ and $v$. More specific, the eigenvectors of $\mathbf{D}$ give the smoothing direction, and the corresponding eigenvalues determine the magnitude of smoothing.

**Homogeneous Regularisation.** First ideas for the smoothness term go back to Horn and Schunck [19] who used a homogeneous regulariser. In our framework it reads as

$$\begin{aligned} V_{\text{H}}(\nabla_2 u, \nabla_2 v) \ &:= \ |\nabla_2 u|^2 + |\nabla_2 v|^2 \\ &= \ u_{\mathbf{s}_1}^2 + u_{\mathbf{s}_2}^2 + v_{\mathbf{s}_1}^2 + v_{\mathbf{s}_2}^2 \ . \end{aligned} \tag{35}$$

The corresponding diffusion tensor is equal to the unit matrix $\mathbf{D}_{\text{H}} = \mathbf{I}$. The smoothing processes thus perform homogeneous diffusion that blurs important flow edges.

**Image-Driven Regularisation.** To obtain sharp flow edges, image-driven methods [1, 27] reduce the smoothing at image edges, indicated by large values of $|\nabla_2 f|^2 = \text{tr } \mathbf{S}_0$.

An isotropic image-driven regulariser goes back to Alvarez et al. [1] who used

$$\begin{aligned} V_{\mathrm{II}}(\nabla_2 u, \nabla_2 v) &:= g\big(|\nabla_2 f|^2\big)\ \big(|\nabla_2 u|^2 + |\nabla_2 v|^2\big) \\ &= g(\mathrm{tr}\ \mathbf{S}_0)\ \big(u_{\mathbf{s}_1}^2 + u_{\mathbf{s}_2}^2 + v_{\mathbf{s}_1}^2 + v_{\mathbf{s}_2}^2\big)\ , \end{aligned} \tag{36}$$

where $g$ is a decreasing, strictly positive weight function. The corresponding diffusion tensor $\mathbf{D}_{\mathrm{II}} = g(\mathrm{tr}\,\mathbf{S}_0)\,\mathbf{I}$ shows that the weight function allows to decrease the smoothing in accordance to the strength of image edges.

The anisotropic image-driven regulariser of Nagel and Enkelmann [27] prevents smoothing of the flow field across image boundaries but encourages smoothing along them. This is achieved by the regulariser

$$V_{\mathrm{AI}}(\nabla_2 u, \nabla_2 v) := \nabla_2^\top u\ \ \mathbf{P}(\nabla_2 f)\ \ \nabla_2 u\ +\ \nabla_2^\top v\ \ \mathbf{P}(\nabla_2 f)\ \ \nabla_2 v\ , \tag{37}$$

where $\mathbf{P}(\nabla_2 f)$ denotes a regularised projection matrix perpendicular to the image gradient. It is defined as

$$\mathbf{P}(\nabla_2 f) := \frac{1}{|\nabla_2 f|^2 + 2\kappa^2} \left(\nabla_2^\perp f \left(\nabla_2^\perp f\right)^\top + \kappa^2 \mathbf{I}\right)\ . \tag{38}$$

with a regularisation parameter $\kappa > 0$. In our common framework, this regulariser can be written as

$$V_{\mathrm{AI}}(\nabla_2 u, \nabla_2 v) = \frac{\kappa^2}{\mathrm{tr}\ \mathbf{S}_0 + 2\kappa^2} \left(u_{\mathbf{s}_1^0}^2 + v_{\mathbf{s}_1^0}^2\right) + \frac{\mathrm{tr}\ \mathbf{S}_0 + \kappa^2}{\mathrm{tr}\ \mathbf{S}_0 + 2\kappa^2} \left(u_{\mathbf{s}_2^0}^2 + v_{\mathbf{s}_2^0}^2\right)\ . \tag{39}$$

The correctness of above rewriting can easily be verified and is based on the observations that $\mathbf{s}_1^0$ and $\mathbf{s}_2^0$ are the eigenvectors of $\mathbf{P}$, and that the factors in front of $(u_{\mathbf{s}_1^0}^2 + v_{\mathbf{s}_1^0}^2)$ and $(u_{\mathbf{s}_2^0}^2 + v_{\mathbf{s}_2^0}^2)$ are the corresponding eigenvalues. The diffusion tensor for the regulariser of Nagel and Enkelmann [27] is identical to the projection matrix: $\mathbf{D}_{\mathrm{AI}} = \mathbf{P}$. Concerning its eigenvectors and eigenvalues, we observe that in the limiting case for $\kappa \to 0$, where $V_{\mathrm{AI}} \to u_{\mathbf{s}_2^0}^2 + v_{\mathbf{s}_2^0}^2$, we obtain a smoothing solely in $\mathbf{s}_2^0$-direction, i.e. along image edges. In the definition of the normal flow (7) we have seen that a data term that models the brightness constancy assumption constraints the flow only orthogonal to image edges. In the limiting case, the regulariser of Nagel and Enkelmann can hence be interpreted as a first complementary smoothness term that fills in information orthogonal to the data constraint direction.

The drawback of image-driven strategies is that they are prone to oversegmentation artefacts in textured image regions where image edges do not necessarily correspond to flow edges.

**Flow-Driven Regularisation.** To remedy the oversegmentation problem, it makes sense to adapt the smoothing process to the *flow edges* instead of the image edges.

In the isotropic setting, Shulman and Hervé [33] and Schnörr [31] proposed to use subquadratic penaliser functions for the smoothness term, i.e.

$$\begin{aligned} V_{\mathrm{IF}}(\nabla_2 u, \nabla_2 v) &:= \Psi_V\big(|\nabla_2 u|^2 + |\nabla_2 v|^2\big) \\ &= \Psi_V\big(u_{\mathbf{s}_1}^2 + u_{\mathbf{s}_2}^2 + v_{\mathbf{s}_1}^2 + v_{\mathbf{s}_2}^2\big) \quad , \end{aligned} \tag{40}$$

where the penaliser function $\Psi_V(s^2)$ is preferably increasing, differentiable and convex in $s$. The associated diffusion tensor is given by

$$\mathbf{D}_{\mathrm{IF}} = \Psi_V'\big(u_{\mathbf{s}_1}^2 + u_{\mathbf{s}_2}^2 + v_{\mathbf{s}_1}^2 + v_{\mathbf{s}_2}^2\big)\ \mathbf{I} \ . \tag{41}$$

The underlying diffusion processes perform nonlinear isotropic diffusion, where the smoothing is reduced at the boundaries of the evolving flow field via the decreasing diffusivity $\Psi_V'$. If one uses the convex penaliser [12]

$$\Psi_V(s^2) := \sqrt{s^2 + \varepsilon^2} \ , \tag{42}$$

one ends up with regularised total variation (TV) regularisation [30] with the diffusivity

$$\Psi_V'(s^2) = \frac{1}{2\sqrt{s^2 + \varepsilon^2}} \approx \frac{1}{2\,|s|} \ . \tag{43}$$

Another possible choice is the non-convex Perona-Malik regulariser (Lorentzian) [6, 29] given by

$$\Psi_V(s^2) := \lambda^2 \log\left(1 + \frac{s^2}{\lambda^2}\right) \quad , \tag{44}$$

that results in Perona-Malik diffusion with the diffusivity

$$\Psi_V'(s^2) = \frac{1}{1 + \frac{s^2}{\lambda^2}} \ , \tag{45}$$

using a contrast parameter $\lambda > 0$.

We will not discuss the anisotropic flow-driven regulariser of Weickert and Schnörr [42] as it does not fit in our framework and also has not been used in the design of our complementary regulariser.

Despite the fact that flow-driven methods reduce the oversegmentation problem caused by image textures, they suffer from another drawback: The flow edges are not as well localised as with image-driven strategies.

**Image- and Flow-Driven Regularisation.** We have seen that image-driven methods suffer from oversegmentation artefacts, but give sharp flow edges. Flow-driven strategies remedy the oversegmentation problem but give less pleasant flow edges. It is thus desirable to combine the advantages of both strategies to obtain sharp flow edges without oversegmentation problems. This aim was achieved by Sun et al. [36] who presented an anisotropic image- and flow-driven smoothness term in a discrete setting. It adapts the smoothing direction to *image* structures but steers the smoothing strength in accordance to the *flow* contrast. In contrast to Nagel and Enkelmann [27] who considered $\nabla_2^\perp f$ to obtain directional information of image structures, the regulariser in [36] analyses the eigenvectors $\mathbf{s}_i$ of the structure tensor $\mathbf{S}_\rho$ from (30) to obtain a more robust direction estimation. A continuous version of this regulariser can be written as

$$V_{\mathrm{AIF}}(\nabla_2 u, \nabla_2 v) := \Psi_V\big(u_{\mathbf{s}_1}^2\big) + \Psi_V\big(v_{\mathbf{s}_1}^2\big) + \Psi_V\big(u_{\mathbf{s}_2}^2\big) + \Psi_V\big(v_{\mathbf{s}_2}^2\big) \quad . \tag{46}$$

Here, we obtain two diffusion tensors, that for $p \in \{u, v\}$ read as

$$\mathbf{D}_{\mathrm{AIF}}^p = \Psi_V'\big(p_{\mathbf{s}_1}^2\big)\ \mathbf{s}_1\,\mathbf{s}_1^\top + \Psi_V'\big(p_{\mathbf{s}_2}^2\big)\ \mathbf{s}_2\,\mathbf{s}_2^\top \quad . \tag{47}$$

We observe that these tensors allow to obtain the desired behaviour: The regularisation direction is adapted to the image structure directions $\mathbf{s}_1$ and $\mathbf{s}_2$, whereas the magnitude of the regularisation depends on the flow contrast encoded in $p_{\mathbf{s}_1}$ and $p_{\mathbf{s}_2}$. As a result, one obtains the same sharp flow edges as image-driven methods but does not suffer from oversegmentation problems.

### 2.2.1 Our Novel Complementary Regulariser

In spite of its sophistication, the anisotropic image- and flow-driven model [36] given in (46) still suffers from a few shortcomings. We introduce three amendments that we will discuss now.

**Regularisation Tensor.** A first remark w.r.t. the model from (46) is that the directional information from the structure tensor $\mathbf{S}_\rho$ is not consistent with the imposed constraints of our data term (27). It is more natural to take into account directional information provided by the motion tensor (19) and to steer the anisotropic regularisation process w.r.t. *"constraint edges"* instead of image edges. To this end we propose to analyse the eigenvectors $\mathbf{r}_1$ and $\mathbf{r}_2$ of the *regularisation tensor*

$$\mathbf{R}_\rho := \sum_{i=1}^{3} K_\rho * \left[\theta_0^i\ \left(\nabla_2 f^i\ \nabla_2^\top f^i\right) + \gamma\left(\theta_x^i\ \left(\nabla_2 f_x^i\ \nabla_2^\top f_x^i\right) + \theta_y^i\ \left(\nabla_2 f_y^i\ \nabla_2^\top f_y^i\right)\right)\right] , \tag{48}$$

which can be regarded as a generalisation of the structure tensor (30). Note that the regularisation tensor differs from the motion tensor $\bar{\mathbf{J}}^c$ from (19) by the facts that it *(i)* integrates neighbourhood information via the Gaussian convolution, and *(ii)* uses the spatial gradient operator $\nabla_2$ instead of the spatio-temporal operator $\nabla_3$. The latter is due to the spatial regularisation. In Section 2.3 we extend our regulariser to the spatio-temporal domain, yielding a regularisation tensor that also uses the spatio-temporal gradient $\nabla_3$. Further note that a Gaussian convolution of the *motion* tensor leads to a combined local-global (CLG) data term in the spirit of [11]. Our experiments in Section 5.1 will analyse in which cases such a modification of our data term can be useful.

**Rotational Invariance.** The smoothness term $V_{\mathrm{AIF}}$ from (46) lacks the desirable property of rotational invariance, because the directional derivatives of $u$ and $v$ in the eigenvector directions are penalised separately. We propose to jointly penalise the directional derivatives, yielding

$$V_{\mathrm{AIF}_{\mathbf{R}_{\rho,\mathrm{RI}}}}(\nabla_2 u, \nabla_2 v) := \Psi_V\Big(u_{\mathbf{r}_1}^2 + v_{\mathbf{r}_1}^2\Big) + \Psi_V\Big(u_{\mathbf{r}_2}^2 + v_{\mathbf{r}_2}^2\Big) \;, \tag{49}$$

where we use the eigenvectors $\mathbf{r}_i$ of the regularisation tensor.

**Single Robust Penalisation.** The above regulariser (49) performs a *twofold robust penalisation* in both eigenvector directions. However, the data term mainly constraints the flow in direction of the largest eigenvalue of the spatial motion tensor, i.e. in $\mathbf{r}_1$-direction. We hence propose a *single robust penalisation* in $\mathbf{r}_1$-direction. In the orthogonal $\mathbf{r}_2$-direction, we opt for a quadratic penalisation to obtain a strong filling-in effect of missing information. The benefits of this design will be confirmed by our experiments in Section 5.2. Incorporating the single robust penalisation finally yields our *complementary regulariser*

$$V_{\mathrm{CR}}(\nabla_2 u, \nabla_2 v) := \Psi_V\Big(u_{\mathbf{r}_1}^2 + v_{\mathbf{r}_1}^2\Big) + u_{\mathbf{r}_2}^2 + v_{\mathbf{r}_2}^2 \;\;, \tag{50}$$

that complements the proposed robust data term from (27) in an optimal fashion. For the penaliser $\Psi_V$, we propose the to use the Perona-Malik regulariser (44).
The corresponding joint diffusion tensor is given by

$$\mathbf{D}_{\mathrm{CR}} = \Psi_V'\big(u_{\mathbf{r}_1}^2 + v_{\mathbf{r}_1}^2\big)\;\; \mathbf{r}_1\mathbf{r}_1^\top + \mathbf{r}_2\mathbf{r}_2^\top \;\;, \tag{51}$$

with $\Psi_V'$ given in (45). The derivation of this diffusion tensor is presented in the Appendix A.

**Discussion.** To understand the advantages of the complementary regulariser compared to the anisotropic image- and flow-driven regulariser (46), we compare our joint diffusion tensor (51) to its counterparts (47), which reveals the following innovations: *(i)* The smoothing direction is adapted to constraint edges instead of image edges, as the eigenvectors of the regularisation tensor $\mathbf{r}_i$ are used instead of the eigenvectors of the structure tensor. *(ii)* We achieve rotational invariance by coupling the two flow components in the argument of $\Psi'_V$. *(iii)* We only reduce the smoothing *across* constraint edges, i.e. in $\mathbf{r}_1$-direction. Along them, always a strong diffusion with strength 1 is performed, resembling edge-enhancing anisotropic diffusion [41].
Furthermore, when analysing our joint diffusion tensor, the benefits of the underlying anisotropic image- and flow-driven regularisation become visible. The smoothing strength across constraint edges is determined by the expression $\Psi'_V(u^2_{\mathbf{r}_1} + v^2_{\mathbf{r}_1})$. Here we can distinguish two scenarios: At a flow edge that corresponds to a constraint edge, the flow gradients will be large and almost parallel to $\mathbf{r}_1$. Thus, the argument of the decreasing function $\Psi'_V$ will be large, yielding a reduced diffusion which preserves this important edge. At "deceiving" texture edges in flat flow regions, however, the flow gradients are small. This results in a small argument for $\Psi'_V$, leading to almost homogeneous diffusion. Hence, we perform a pronounced smoothing in both directions that avoids oversegmentation artefacts.
Finally note that our complementary regulariser has the same structure, even if other data terms are used. Only the regularisation tensor $\mathbf{R}_\rho$ has to be adapted to the new data term.

### 2.2.2 Summary

To conclude this section, Table 1 summarises the discussed regularisers rewritten in our framework. It also compares the way directional information is obtained for anisotropic strategies, and it indicates if the regulariser is rotationally invariant. Note that despite the fact these regularisers have been developed within almost three decades, our taxonomy shows their structural similarities.

## 2.3 Extension to a Spatio-Temporal Smoothness Term

The smoothness terms we have discussed so far model the assumption of a *spatially* smooth flow field. As image sequences in general encompass more than two frames, yielding several flow fields, it makes sense to also assume a temporal smoothness of the flow fields, leading to spatio-temporal regularisation strategies.

Table 1: Comparison of regularisation strategies. The next to last column names the tensor that is analysed to obtain directional information for anisotropic strategies, and the last column indicates if the corresponding regulariser is rotationally invariant.

| Strategy | Regulariser $V$ | Directional Adaptation | Rotationally Invariant |
|---|---|---|---|
| Homogeneous [19] | $u_{\mathbf{s}_1}^2 + u_{\mathbf{s}_2}^2 + v_{\mathbf{s}_1}^2 + v_{\mathbf{s}_2}^2$ | — | ✓ |
| Isotropic image-driven [1] | $g(\mathrm{tr}\ \mathbf{S}_0)\ \left(u_{\mathbf{s}_1}^2 + u_{\mathbf{s}_2}^2 + v_{\mathbf{s}_1}^2 + v_{\mathbf{s}_2}^2\right)$ | — | ✓ |
| Anisotropic image-driven [27] | $u_{\mathbf{s}_2^0}^2 + v_{\mathbf{s}_2^0}^2$, for $\kappa \to 0$ | $\mathbf{S}_0$ | ✓ |
| Isotropic flow-driven [33, 31] | $\Psi_V\left(u_{\mathbf{s}_1}^2 + u_{\mathbf{s}_2}^2 + v_{\mathbf{s}_1}^2 + v_{\mathbf{s}_2}^2\right)$ | — | ✓ |
| Anisotropic image and flow-driven [36] | $\Psi_V\left(u_{\mathbf{s}_1}^2\right) + \Psi_V\left(v_{\mathbf{s}_1}^2\right) + \Psi_V\left(u_{\mathbf{s}_2}^2\right) + \Psi_V\left(v_{\mathbf{s}_2}^2\right)$ | $\mathbf{S}_\rho$ | — |
| **Anisotropic complementary image- and flow-driven** | $\Psi_V\left(u_{\mathbf{r}_1}^2 + v_{\mathbf{r}_1}^2\right) + u_{\mathbf{r}_2}^2 + v_{\mathbf{r}_2}^2$ | $\mathbf{R}_\rho$ | ✓ |

A spatio-temporal (ST) version of the general energy functional (1) reads as

$$E^{\mathrm{ST}}(u,v) = \int\limits_{\Omega\times[0,T]} \left[M(u,v) + \alpha\, V^{\mathrm{ST}}(\nabla_3 u, \nabla_3 v)\right] \,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t \ . \tag{52}$$

Compared to the spatial energy (1) we note the additional integration over the time domain and that the smoothness term now depends on the spatio-temporal flow gradient.
To extend our complementary regulariser from (50) to the spatio-temporal domain, we define the *spatio-temporal regularisation tensor*

$$\mathbf{R}_\rho^{\mathrm{ST}} := K_\rho * \bar{\mathbf{J}}^c \ . \tag{53}$$

For $\rho = 0$ it is identical to the motion tensor $\bar{\mathbf{J}}^c$ from (19). The Gaussian convolution with $K_\rho$ is now performed in the spatio-temporal domain, which also holds for the presmoothing of the image sequence. The spatio-temporal regularisation tensor is a $3\times 3$ tensor that possesses three orthonormal eigenvectors $\mathbf{r}_1, \mathbf{r}_2$ and $\mathbf{r}_3$. With their help, we define the *spatio-temporal complementary regulariser (ST-CR)*

$$V_{\mathrm{CR}}^{\mathrm{ST}}(\nabla_3 u, \nabla_3 v) := \Psi_V\Big(u_{\mathbf{r}_1}^2 + v_{\mathbf{r}_1}^2\Big) + u_{\mathbf{r}_2}^2 + v_{\mathbf{r}_2}^2 + u_{\mathbf{r}_3}^2 + v_{\mathbf{r}_3}^2 \ . \tag{54}$$

The corresponding spatio-temporal diffusion tensor reads as

$$\mathbf{D}_{\mathrm{CR}}^{\mathrm{ST}} = \Psi_V'\big(u_{\mathbf{r}_1}^2 + v_{\mathbf{r}_1}^2\big)\ \mathbf{r}_1\,\mathbf{r}_1^\top + \mathbf{r}_2\,\mathbf{r}_2^\top + \mathbf{r}_3\,\mathbf{r}_3^\top \ . \tag{55}$$

# 3 Automatic Selection of the Smoothness Weight

The last step missing for our OFH method is a strategy that automatically determines the optimal smoothness parameter $\alpha$ for the image sequence under consideration. This is especially important in real world applications of optic flow where no ground truth flow is known. Note that if the latter would be the case, we could simply select the smoothness weight that gives the flow field with the smallest deviation from the ground truth.

## 3.1 A Novel Concept

We propose an error measure that allows to judge the quality of a flow field without knowing the ground truth. This error measure bases on a novel

concept, the optimal prediction principle (OPP). The OPP states that the flow field obtained with an optimal smoothness weight allows for the best prediction of the next frames in the image sequence. This makes sense as a too small smoothness weight would lead to an overfit to the first two frames and consequently result in a bad prediction of further frames. For too large smoothness weights, the flow fields would be too smooth and thus also lead to a bad prediction.
Following the OPP, our error measure needs to judge the quality of the prediction achieved with a given flow field. To this end, we evaluate the imposed data constraints between the first and the third frame of the image sequence, resulting in an *average data constancy error (ADCE)* measure. To compute this measure, we assume that the motion of the scene objects is of more or less constant speed and that it describes linear trajectories within the considered three frames. Under these assumptions, we simply double the flow vectors to evaluate the data constraints between first and third frame. Following this strategy, we can define the ADCE between frame 1 and 3 as

$$\begin{aligned} \mathrm{ADCE}_{1,3}(\mathbf{w}_\alpha) := \frac{1}{|\Omega|} \int_\Omega \Bigg[ \sum_{i=1}^{3} \Psi_M\Big(\theta_0^i \left(f^i(\mathbf{x}+2\mathbf{w}_\alpha) - f^i(\mathbf{x})\right)^2\Big) \\ + \gamma \left( \sum_{i=1}^{3} \Psi_M\Big(\theta_x^i \left(f_x^i(\mathbf{x}+2\mathbf{w}_\alpha) - f_x^i(\mathbf{x})\right)^2 \right. \\ \left. + \theta_y^i \left(f_y^i(\mathbf{x}+2\mathbf{w}_\alpha) - f_y^i(\mathbf{x})\right)^2 \Big) \right) \Bigg] \,\mathrm{d}x\,\mathrm{d}y \ , \end{aligned} \tag{56}$$

where $\mathbf{w}_\alpha$ denotes the flow field obtained with a smoothness weight $\alpha$. The integrand of above expression is (apart from the doubled flow field) a variant of our final data term (27) where no linearisation of the constancy assumptions have been performed. To evaluate the images at the subpixel locations $f^i(\mathbf{x}+2\mathbf{w}_\alpha)$ we use Coons patches based on bicubic interpolation [13].

## 3.2 Determining the Best Parameter

In general, the relation between $\alpha$ and the ADCE is not convex, which excludes the use of gradient descent-like approaches for finding the optimal value of $\alpha$ w.r.t. our error measure.
We propose a brute-force method similar to the one of Ng and Solo [28]: We first compute the error measures for a "sufficiently large" set of flow fields obtained with different $\alpha$ values. We then select the $\alpha$ that gives the smallest error. To reduce the number of $\alpha$ values to test, we propose to start from a given, standard value $\alpha_0$, say. This value is then incremented/decremented

$n_\alpha$ times by multiplying/dividing it with a stepping factor $a > 1$, yielding in total $2\,n_\alpha + 1$ tests. This strategy results in testing more values of $\alpha$ that are close to $\alpha_0$ and, more important, tests less very small or very large values of $\alpha$ that hardly give reasonable results.

# 4 Implementation

The solution of the Euler-Lagrange equations for our method comes down to solving a nonlinear system of equations. We solve the system by a nonlinear multigrid scheme based on a Gauß-Seidel type solver with alternating line relaxation [10].

## 4.1 Warping Strategy for Large Displacements

The derivation of the optic flow constraint (4) by means of a linearisation is only valid under the assumption of small displacements. If the temporal sampling of the image sequence is too coarse, this precondition will be violated and a linearised approach fails. To overcome this problem, Brox et al. [8] proposed a coarse-to-fine multiscale warping strategy. To obtain a coarse representation of the problem, we downsample the input images by a factor $\eta \in [0.5, 1.0)$. Prior to downsampling, we apply a low-pass filter to the images by performing a Gaussian convolution with standard deviation $\sqrt{2}/(4\eta)$. This prevents aliasing problems.
At each warping level, we split the flow field into an already computed solution from coarser levels and an unknown flow increment. As the increments are small, they can computed by the presented linearised approach. At the next finer level, the already computed solution serves as initialisation, which is achieved by performing a motion compensation of the second frame by the current flow, known as warping. For warping with subpixel precision we again use Coons patches based on bicubic interpolation [13].

**Adapting the Smoothness Weight to the Warping Level.** The influence of the data term usually becomes smaller at coarser levels of our multiscale framework. This is due to the smoothing properties of the downsampling that leads to smaller values of the image gradients at coarse levels. Such a behaviour is in fact desirable as the data term might not be reliable at coarse levels. Our proposed data term normalisation leads, however, to image gradients that are approximately in the same range at each level. To recover the previous reduction of the data term at coarse levels, we propose to adapt the smoothness weight $\alpha$ to the warping level $k$. This is achieved

by setting $\alpha(k) = \alpha/\eta^k$ which results in larger values of $\alpha$ and an emphasis of the smoothness term at coarse levels.

### 4.2 Discretisation

We follow [10] for the discretisation of the Euler-Lagrange equations. The images and the flow fields are sampled on a rectangular pixel grid with grid size $h$ and temporal step size $\tau$.
Spatial image derivatives are approximated via central finite differences using the stencil $\frac{1}{12h}(1, -8, 0, 8, -1)$, resulting in a fourth order approximation. The spatial flow derivatives are discretised by second order approximations with the stencil $\frac{1}{2h}(-1, 0, 1)$. For approximating temporal image derivatives we use a two-point stencil $(-1, 1)$, resulting in a temporal difference. Concerning the temporal flow derivatives that occur in the spatio-temporal case, we use the stencil $(-1, 1)/\tau$. Here, it makes sense to adapt the value of $\tau$ to the given image sequence to allow for an appropriate scaling of the temporal direction compared to the spatial directions [43].
When computing the motion tensor, occurring derivatives are averaged from the two frames at time $t$ and $t + 1$ to obtain a lower approximation error. For the regularisation tensor, the derivatives are solely computed at the first frame as we only want to consider directional information from the reference image.

## 5 Experiments

In our experiments we show the benefits of the OFH approach. The first experiments are concerned with our robust data term and the complementary smoothness term in the spatial and the spatio-temporal domain. Then, we turn to the automatic selection of the smoothness weight. After a small experiment on the importance of anti-aliasing in the warping scheme, we finish our experiments by presenting the performance at the Middlebury optic flow benchmark [2] (*http://vision.middlebury.edu/flow/eval/*).
As all considered sequences exhibit relatively large displacements, we use the multiscale warping approach described in Section 4.1. The flow fields are visualised by a colour code where hue encodes the flow direction and brightness the magnitude, see Figure 3 (d). Throughout our experiments we use constant values for the following parameters: $\zeta = 0.1, \varepsilon = 0.001, \lambda = 0.1$.

## 5.1 Robust Data Term

**Benefits of Normalisation and the HSV Colour Space.** We proposed two main innovations in the data term: constraint normalisation and using an HSV colour representation. In our first experiment, we thus compare our method against variants *(i)* without data term normalisation. *(ii)* using the RGB instead of the HSV colour space. For the latter we only separately robustify the brightness and the gradient constancy assumption, as a separate robustification of the RGB channels makes no sense. In Figure 3 we show the results for the *Snail* sequence that we have created. Note that it is a rather challenging sequence due to severe shadows and large displacements up to 25 pixels. When comparing the results to our result in Figure 3 (i), the following drawbacks of the modified versions become obvious: Without data term normalisation (Figure 3 (e)), unpleasant artefacts at image edges arise, even when using a large smoothness weight $\alpha$. When relying on the RGB colour space (Figure 3 (f)), a phantom motion in the shadow region at the right border is estimated.

**Effect of the Separate Robust Penalisation.** This experiment illustrates the desirable effect of our separate robust penalisation of the HSV channels. Using the *Rubberwhale* sequence from the Middlebury database, we show in Figure 4 the data term weights $\Psi'_M(\mathbf{w}^\top \overline{\mathbf{J}}_0^i \mathbf{w})$ for the brightness constancy assumption on the hue, the saturation and the value channel ($i = 1, .., 3$). Here, brighter pixels correspond to a larger weight and we only show a zoom for better visibility. As we can see, the weight of the value channel is reduced in the shadow regions (left of the wheel, of the orange toy and of the clam). This is desirable as the value channel is not invariant under shadows, see Figure 2.

**A CLG Variant of Our Method.** Our next experiment is concerned with a CLG variant of our data term where we, as for the regularisation tensor, perform a Gaussian convolution of the motion tensor entries.
First, we compare our method against a CLG variant for some Middlebury sequences, see Table 2. To evaluate the quality of the flow fields compared to the given ground truth, we use the *average angular error (AAE)* measure ([3]). We find that the CLG variant always leads to worse results and conclude that for the considered test sequences, this modification seems not to be useful.
The flow fields for our proposed method are visualised in Figure 5. Together with the parameter settings, the resulting error measures can be found in the caption of the figure. To ease comparison with other methods, we give the

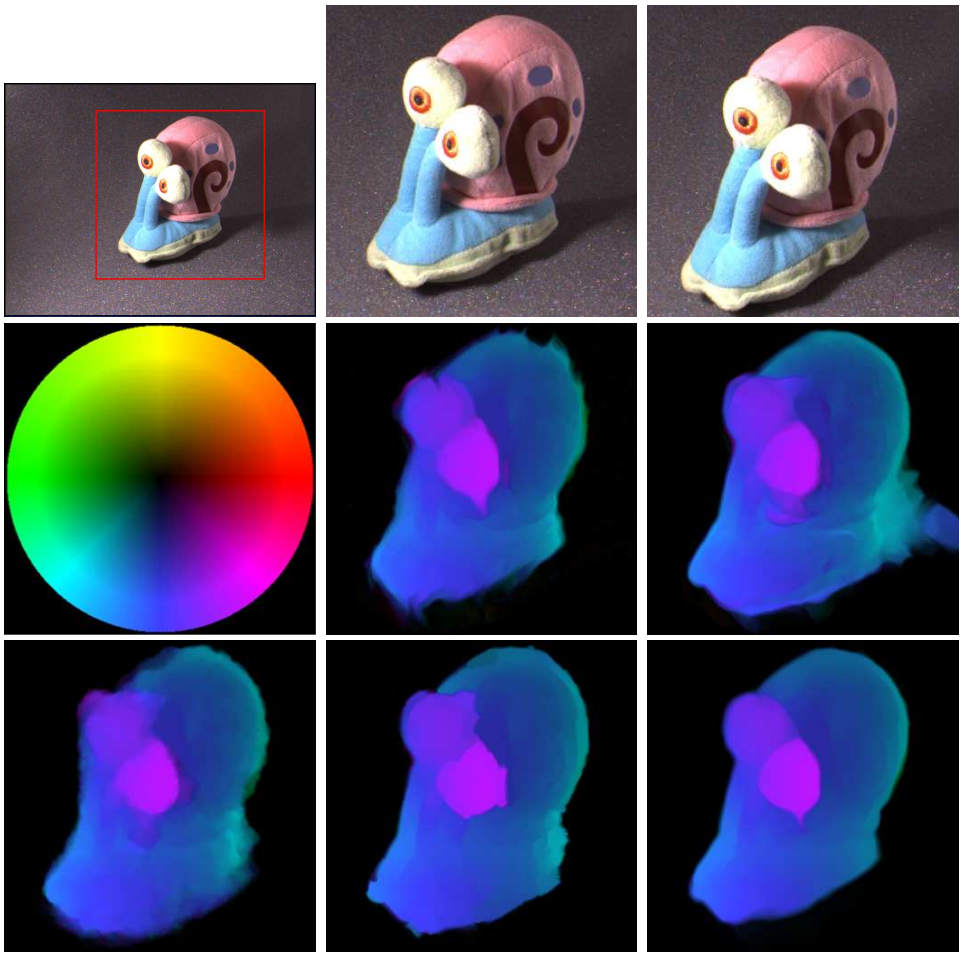

Figure 3: Results for our *Snail sequence* with different variants of our method. *First row, from left to right:* **(a)** First frame. **(b)** Zoom in marked region of first frame. **(c)** Same for second frame. *Second row, from left to right:* **(d)** Colour code. **(e)** Flow field in marked region, without normalisation ($\alpha = 5000.0$). **(f)** Same for RGB colour space ($\alpha = 300.0$). *Third row, from left to right:* **(g)** Same for TV regularisation ($\alpha = 50.0$). **(h)** Same for image- and flow-driven regularisation [36] ($\alpha = 2000.0$). **(i)** Same for our method ($\alpha = 2000.0$). All results used the fixed parameters $\sigma = 0.5, \gamma = 20.0, \rho = 4.0, \eta = 0.95$

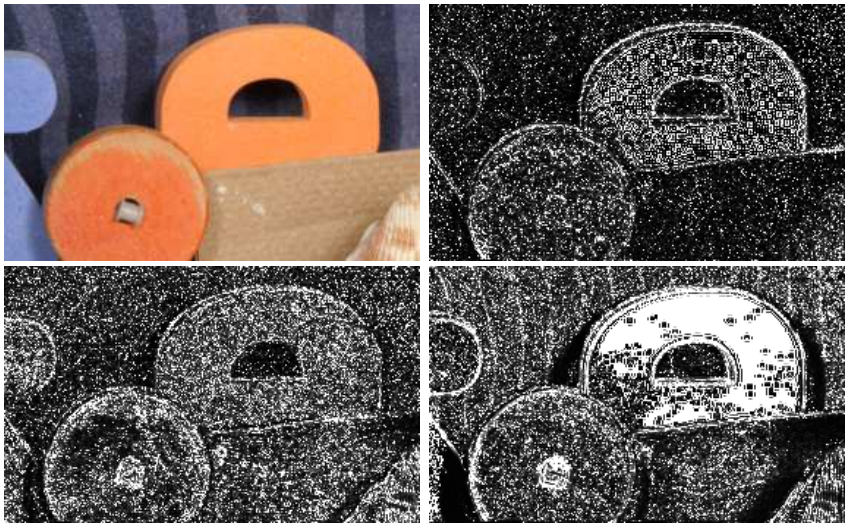

Figure 4: Effect of our separate robust penalisation of the HSV channels. *First row, from left to right:* **(a)** Zoom in first frame of the *Rubberwhale* sequence. **(b)** Visualisation of the corresponding hue channel weight. Brighter pixels correspond to a larger weight. *Second row, from left to right:* **(c)** Same for the saturation channel. **(d)** Same for the value channel

AAE and also the alternative *average endpoint error (AEE)* measure ([2]). Concerning the the *Rubberwhale* sequence in the first row of Figure 5, we wish to note that using the HSV colour space prevents unpleasant artefacts in the shadow regions, e.g. left of the wheel.
Although we have seen that a CLG variant of our method does not improve the results on the Middlebury data set, this modification can actually be useful in a certain scenario, namely in the presence of severe noise in the image sequence. To prove this, we compare in Table 3 the performance of our method to its CLG counterpart on noisy versions of the *Yosemite* sequence. As it turns out, the CLG variant improves the results at large noise scales, but deteriorates the quality for low noise scenarios. This also explains the experienced behaviour on the Middlebury data sets, which hardly suffer from noise.

## 5.2 Complementary Smoothness Term

**Comparison with Other Regularisers.** In Figure 3, we compare our method against two results obtained when using another regulariser in conjunction with our robust data term: *(i)* Using the popular TV regulariser; see (40) and (42). *(ii)* Using the anisotropic image and flow-driven regulariser

Table 2: Comparison of our method to a CLG variant (using the AAE)

| Sequence | *Rubberwhale* | *Dimetrodon* | *Grove2* | *Urban2* |
|---|---|---|---|---|
| CLG | 3.00° | 1.59° | 2.20° | 2.67° |
| **Proposed** | 2.77° | 1.54° | 2.16° | 2.49° |

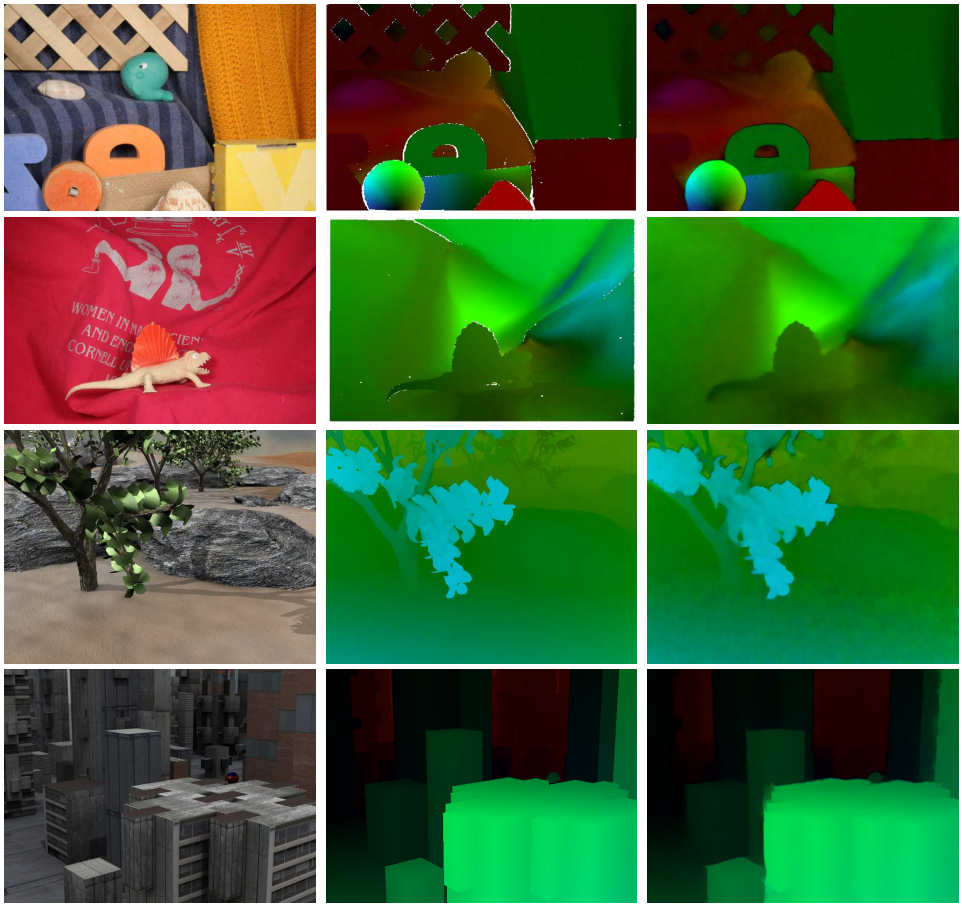

Figure 5: Results for some Middlebury sequences with ground truth. *First column:* Reference frame. *Second column:* Ground truth (white pixels mark locations where no ground truth is given). *Third column:* Result with our method. *From top to bottom: Rubberwhale* ($\alpha = 850.0, \sigma = 0.3, \gamma = 20.0, \rho = 2.0 \Longrightarrow$ AAE $= 2.77°$, AEE $= 0.083$), *Dimetrodon* ($\alpha = 2500.0, \sigma = 0.7, \gamma = 25.0, \rho = 2.0 \Longrightarrow$ AAE $= 1.54°$, AEE $= 0.079$), *Grove2* ($\alpha = 35.0, \sigma = 0.5, \gamma = 0.2, \rho = 1.0 \Longrightarrow$ AAE $= 2.16°$, AEE $= 0.151$), and *Urban2* ($\alpha = 125.0, \sigma = 0.5, \gamma = 1.0, \rho = 1.5 \Longrightarrow$ AAE $= 2.49°$, AEE $= 0.245$). The downsampling rate was set to $\eta = 0.95$ for all sequences

Table 3: Comparison of our method to a CLG variant on noisy versions of the *Yosemite* sequence (using the AAE). We have added Gaussian noise with zero mean and standard deviation $\sigma_n$

| $\sigma_n$ | 0 | 10 | 20 | 40 |
|---|---|---|---|---|
| CLG | 1.75° | 4.01° | **5.82°** | **8.14°** |
| Proposed | **1.64°** | **3.82°** | 6.08° | 8.55° |

from [36], which built the basis of our complementary regulariser. Here, we use a rotationally invariant formulation that can be obtained from (49) by replacing the eigenvectors $\mathbf{r}_i$ of the regularisation tensor by the eigenvectors $\mathbf{s}_i$ of the structure tensor. Comparing the obtained results to our result in Figure 3 (i), we see that TV regularisation (Figure 3 (g)), leads to blurred and badly localised flow edges. Using the regulariser from [36] (Figure 3 (h)), unpleasant staircasing artefacts deteriorate the result.

**Optic Flow in the Spatio-Temporal Domain.** Let us now turn to the spatio-temporal extension of our complementary smoothness term. As most Middlebury sequences consist of 8 frames, a spatio-temporal method would in general be applicable. However, the displacements between two subsequent frames are often rather large there, resulting in a violation of the assumption of a temporally smooth flow field. Consequently, spatio-temporal methods do not improve the results. In our experiments, we use the *Marble* sequence (available at *http://i21www.ira.uka.de/image_sequences/*) and the *Yosemite* sequence from the Middlebury datasets. These sequences exhibit relatively small displacements and our spatio-temporal method allows to obtain notably better results, see Figure 6 and Table 4–5. Note that when using more than two frames, a smaller smoothness weight $\alpha$ has to be chosen and that a too large temporal window may also deteriorate the results again.

## 5.3 Automatic Selection of the Smoothness Weight

**Performance of our Proposed Error Measure.** We first show that our proposed data constancy error between frame 1 and 3 ($\text{ADCE}_{1,3}$) is a very good approximation of the popular angular error (AAE) measure. To this end, we compare the two error measures for the *Grove2* sequence in Figure 7. It becomes obvious that our proposed error measure (Figure 7 (b)) indeed exhibits a shape very close to the angular error shown in Figure 7 (a). As

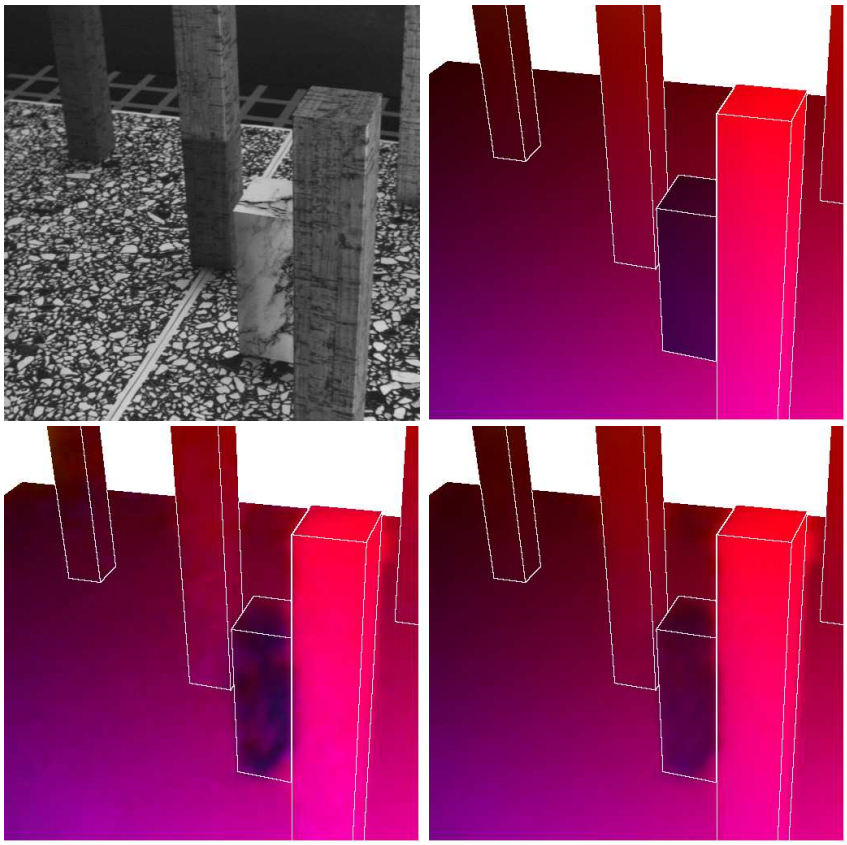

Figure 6: Results for the *Marble* sequence with our spatio-temporal method. *First row, from left to right:* **(a)** Reference frame (frame 16). **(b)** Ground truth (white pixels mark locations where no ground truth is available). *Second row, from left to right:* **(c)** Result using 2 frames (16–17). **(d)** Same for 6 frames (14–19)

Table 4: Smoothness weight $\alpha$ and AAE measures for our spatio-temporal method on the *Marble* sequence, see Figure 6. All results used the fixed parameters $\sigma = 0.5, \gamma = 0.5, \rho = 1.0, \tau = 1.5, \eta = 0.5$. When using more than two frames, the convolutions with $K_\sigma$ and $K_\rho$ are performed in the spatio-temporal domain

| Number of frames (from – to) | 2 (16–17) | 4 (15–18) | 6 (14–19) | 8 (13–20) |
|---|---|---|---|---|
| Smoothn. weight $\alpha$ | 75.0 | 50.0 | 50.0 | 50.0 |
| AAE | 4.85° | 2.63° | **1.86°** | 2.04° |

Table 5: Smoothness weight $\alpha$ and AAE measures for our spatio-temporal method on the *Yosemite* sequence from Middlebury. All results used the fixed parameters $\sigma = 1.0, \gamma = 20.0$, $\rho = 1.5, \tau = 1.0, \eta = 0.5$. Here, better results could be obtained when disabling the temporal presmoothing

| Number of frames (from – to) | 2 (10–11) | 4 (9–12) | 6 (8–13) | 8 (7–14) |
|---|---|---|---|---|
| Smoothn. weight $\alpha$ | 2000.0 | 1000.0 | 1000.0 | 1000.0 |
| AAE | 1.65° | 1.16° | 1.05° | **1.01°** |

our error measure reflects the quality of the prediction with the given flow field, our result further substantiate the validity of the proposed OPP.

**Benefits of an Automatic Parameter Selection.** Next, we show that our automatic parameter selection works well for a large variety of different test sequences. In Table 6, we summarise the AAE obtained when *(i)* setting $\alpha$ to a fixed value ($\alpha = \alpha_0 = 400.0$), *(ii)* using our automatic parameter selection method, and *(iii)* selecting the (w.r.t. the AAE) optimal value of $\alpha$ under the tested proposals. As we can see, estimating $\alpha$ with our proposed method allows to improve the results compared to a fixed value of $\alpha$ in almost all cases. Just for the *Grove 3* sequence, the fixed value of $\alpha$ by accidentally coincides with the optimal value. Compared to the results achieved with an optimal value of $\alpha$, our results are on average 3% and at most 10% worse than the optimal result.

## 5.4 Importance of Anti-Aliasing in the Warping Scheme

We proposed to presmooth the images prior to downsampling in order to avoid aliasing problems. In most cases, the resulting artefacts will not significantly deteriorate the flow estimation, which can be attributed to the robust data term. However, for the *Urban* sequence from the official Middlebury benchmark, anti-aliasing is crucial for obtaining reasonable results, see Figure 8. As it turns out, the large displacement of the building in the lower left corner can only be estimated when using with anti-aliasing. We explain this by the high frequent stripe pattern on the facade of the building.

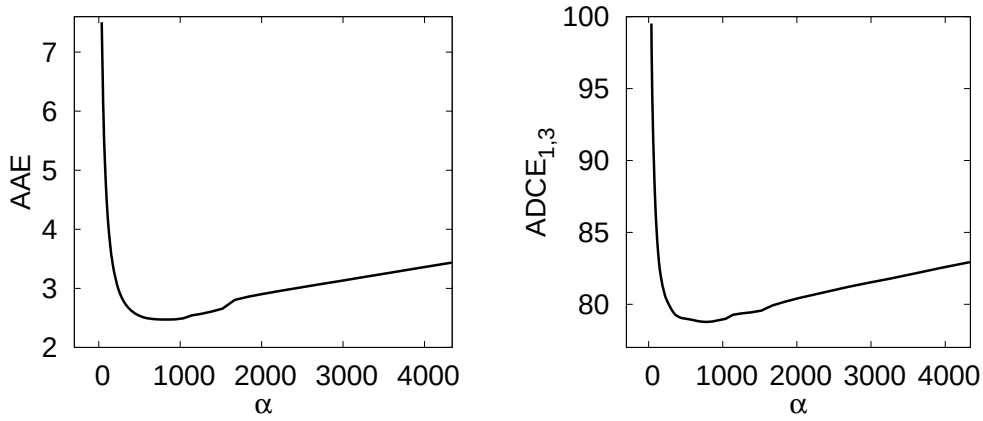


Figure 7: Automatic selection of the smoothness weight $\alpha$ at the *Grove2* sequence. *From left to right:* **(a)** Angular error (AAE) for 51 values of $\alpha$, computed from $\alpha_0 = 400.0$ and a stepping factor $a = 1.1$. **(b)** Same for the proposed data constancy error ($\text{ADCE}_{1,3}$). Remaining parameters were set fixed to $\sigma = 0.5, \gamma = 20.0, \rho = 4.0, \eta = 0.95$

Table 6: Results (AAE) for some Middlebury sequences when *(i)* fixing the smoothness weight ($\alpha = 400.0$), *(ii)* estimating $\alpha$, and *(iii)* with the optimal value of $\alpha$

| Sequence | Fixed $\alpha$ | Estimated $\alpha$ | Optimal $\alpha$ |
|---|---|---|---|
| *Rubberwhale* | 3.43° | 3.00° | 3.00° |
| *Grove2* | 2.59° | 2.43° | 2.43° |
| *Grove3* | 5.50° | 5.62° | 5.50° |
| *Urban2* | 3.22° | 2.84° | 2.66° |
| *Urban3* | 3.44° | 3.37° | 3.35° |
| *Hydrangea* | 1.96° | 1.94° | 1.86° |
| *Yosemite* | 2.56° | 1.89° | 1.71° |
| *Marble* | 5.73° | 5.05° | 4.94° |

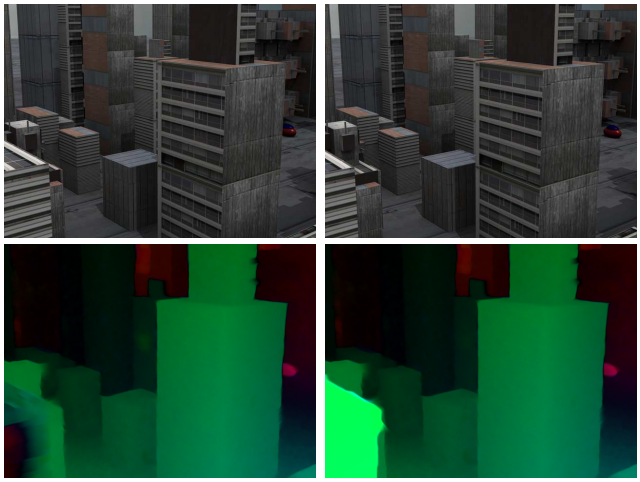

Figure 8: Importance of anti-aliasing on the example of the *Urban* sequence. *Top row, from left to right:* **(a)** Frame 10. **(b)** Frame 11. *Bottom row, from left to right:* **(c)** Our result without anti-aliasing. **(d)** Same with anti-aliasing. All results used the fixed parameters $\alpha = 500.0, \sigma = 0.5, \gamma = 20.0, \rho = 4.0, \eta = 0.95$. Please note that for this sequence, no ground truth is publicly available

## 5.5 Comparison to State-of-the-Art Methods

To compare our method to the state-of-the-art in optic flow estimation, we submitted our results to the popular Middlebury benchmark (available at *http://vision.middlebury.edu/flow/eval/*).
We found that for the provided benchmark sequences, using a HSV colour representation is not as beneficial as seen in our experiment from Figure 3. As the Middlebury sequences hardly suffer from difficult illumination conditions, we cannot profit from the photometric invariances of the HSV colour space. On the other hand, some sequences even pose problems in their HSV representation. As an example, consider the results for the the *Teddy* sequence in the first row of Figure 9. Here we see that the small white triangle beneath the chimney causes unpleasant artefacts in the flow field. This results from the problem that greyscales do not have a unique representation in the hue as well as the saturation channel. Nevertheless, there are also sequences where a HSV colour representation is beneficial. For the *Mequon* sequence (second row of Figure 9) a HSV colour representation removes artefacts in the shadows left of the toys. The bottom line is, however, that for the whole set of benchmark sequences, we obtain slightly better results when using the RGB colour space. Thus, we use this variant of our method for

Table 7: Estimated values of the smoothness weight $\alpha$, using our automatic parameter selection method

| Sequence | Smoothness weight $\alpha$ |
|---|---|
| *Army* | 277.8 |
| *Grove* | 277.8 |
| *Mequon* | 277.8 |
| *Schefflera* | 691.2 |
| *Urban* | 480.0 |
| *Wooden* | 995.3 |
| *Yosemite* | 1433.3 |

evaluation at the Middlebury benchmark.
For our submission we used in accordance to the guidelines a fixed set of parameters: $\sigma = 0.5, \gamma = 20.0, \rho = 4.0, \eta = 0.95$. The smoothness weight $\alpha$ was automatically determined by our proposed method with the settings $n_\alpha = 8, \alpha_0 = 400.0, a = 1.2$. For the *Teddy* with only two frames, we set $\alpha = \alpha_0$, as our parameter estimation method is not applicable in this case. The resulting running time for the *Urban* sequence (640×480 pixels) was 620 s on a standard PC (3.2 GHz Intel Pentium 4). For the parameter selection we computed $2 \cdot 8 + 1 = 17$ flow fields, corresponding to approximately 36 s per flow field. As recently shown by Gwosdek et al. [18], these runtimes can be significantly lowered by a parallel implementation on modern GPUs.
At the time of submission (August 2010), we achieve the 4th place w.r.t. the AAE and the AEE measure among 39 listed methods. Note that our previous Complementary Optic Flow method [46] only ranks 6th for the AAE and 9th for the AEE, which demonstrates the benefits of the proposed novelties in this paper, like the automatic parameter selection and the anti-aliasing.
In Table 7 we additionally summarise the estimated values of $\alpha$ resulting from our automatic parameter selection method. As desired, for sequences with small details in the flow field (*Army*, *Grove*, *Mequon*) a small smoothness weight is chosen. On the other hand, sequences like *Wooden* and *Yosemite* with a rather smooth flow yield significantly larger values for the smoothness weight.

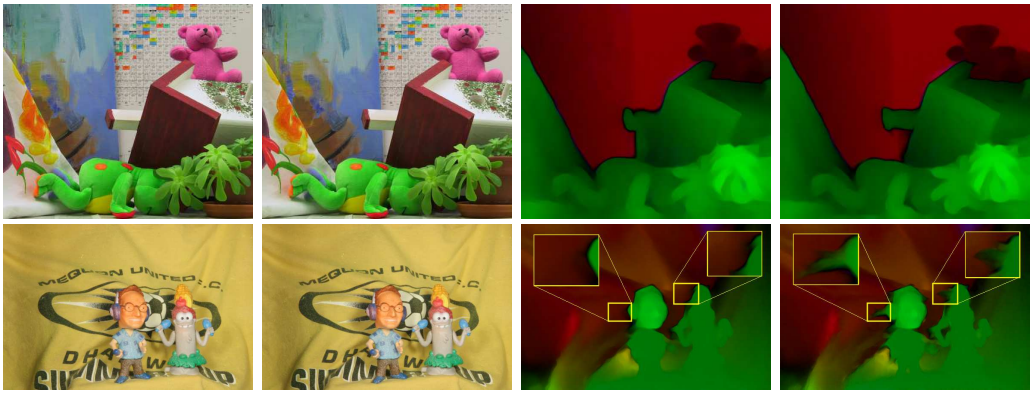

Figure 9: Comparison of results obtained with HSV or RGB colour representation. *First row, from left to right:* **(a)** Frame 10 of the *Teddy* sequence. **(b)** Frame 11. **(c)** Result when using the HSV colour space (AAE = 3.94°). **(d)** Same for the RGB colour space (AAE = 2.64°). *Second row, from left to right:* **(e)** Frame 10 of the *Mequon* sequence. **(f)** Frame 11. **(g)** Result when using the HSV colour space (AAE = 2.28°). **(h)** Same for the RGB colour space (AAE = 2.84°)

# 6 Conclusions and Outlook

In this paper we have shown how to harmonise the three main constituents of variational optic flow approaches: the data term, the smoothness term and the smoothness weight. This was achieved by two main ideas: *(i)* We developed a smoothness term that achieves an optimal complementary smoothing behaviour w.r.t. the imposed data constraints. *(ii)* We presented a simple, yet well performing method for determining the optimal smoothness weight for the given the image sequence. To this end, we came up with a novel paradigm, the optimal prediction principle (OPP).

Our optic flow in harmony (OFH) method bases on an advanced data term that combines and extended successful concepts like normalisation, photometric invariant colour representation, higher order constancy assumptions and robust penalisation. The anisotropic complementary smoothness term incorporates directional information from the motion tensor occurring in the data term. The smoothing in data constraint direction is reduced to avoid interference with the data term, while a strong smoothing in the orthogonal direction allows to fill-in missing information. This yields an optimal complementary between both terms. Furthermore, our smoothness term unifies the benefits of image- and flow-driven regularisers, resulting in sharp flow edges without oversegmentation artefacts. The proposed parameter selec-

tion method bases on the OPP that we introduced in this paper. It states the flow field obtained with an optimal smoothness weight allows for the best prediction of the next frames in the image sequence. Under mild assumptions, the quality of the prediction can be judged by evaluating the data constraints between first and third frame of the sequence and using the doubled flow vectors. Due to its simplicity, our method can easily be used in all variational optic flow approaches and additionally gives surprisingly good results.
The benefits of the OFH idea are demonstrated by our extensive experimental validation and the competitive performance at the Middlebury optic flow benchmark. Our paper thus shows that a careful design of data and smoothness term together with an automatic choice of the smoothness weight allows to outperform other well-engineered methods that incorporate many more processing steps, e.g. segmentation [22], or the integration of an epipolar geometry prior [39].
We hope that our work will give rise to more "harmonised" approaches in other fields where energy-based methods are used, e.g. image registration. Our current research is concerned with exploring further improvements in the data and smoothness term. For the latter, incorporating recent non-local smoothing strategies [35, 44] into our framework can be interesting. Finally, further investigations on the presented novel parameter selection approach seem promising.

# Acknowledgements

Henning Zimmer gratefully acknowledges funding by the International Max-Planck Research School (IMPRS).

# References

[1] L. Alvarez, J. Esclarín, M. Lefébure, and J. Sánchez. A PDE model for computing the optical flow. In *Proc. XVI Congreso de Ecuaciones Diferenciales y Aplicaciones*, pages 1349–1356, Las Palmas de Gran Canaria, Spain, September 1999.

[2] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. Technical Report MSR-TR-2009-179, Microsoft Research, Redmond, WA, December 2009.

[3] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, February 1994.

[4] M. Bertero, T. A. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889, August 1988.

[5] J. Bigün, G. H. Granlund, and J. Wiklund. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):775–790, August 1991.

[6] 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, January 1996.

[7] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, MA, 1987.

[8] 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.

[9] 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, October 2005. IEEE Computer Society Press.

[10] 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, December 2006.

[11] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211–231, 2005.

[12] I. Cohen. Nonlinear variational method for optical flow computation. In *Proc. Eighth Scandinavian Conference on Image Analysis*, volume 1, pages 523–530, Tromsø, Norway, May 1993.

[13] S. A. Coons. Surfaces for computer aided design of space forms. Technical Report MIT/LCS/TR-41, Massachusetts Institute of Technology, Cambridge, MA, June 1967.

[14] L. E. Elsgolc. *Calculus of Variations*. Pergamon, London, England, 1962.

[15] D. J. Fleet and A. D. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5(1):77–104, August 1990.

[16] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, Interlaken, Switzerland, June 1987.

[17] P. Golland and A. M. Bruckstein. Motion from color. *Computer Vision and Image Understanding*, 68(3):346–362, December 1997.

[18] P. Gwosdek, H. Zimmer, S. Grewenig, A. Bruhn, and J. Weickert. A highly efficient GPU implementation for variational optic flow based on the Euler-Lagrange framework. In *Proc. 2010 ECCV Workshop on Computer Vision with GPUs*, Heraklion, Greece, September 2010.

[19] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[20] K. Krajsek and R. Mester. Bayesian model selection for optical flow estimation. In F. A. Hamprecht, C. Schnörr, and B. Jähne, editors, *Pattern Recognition*, volume 4713 of *Lecture Notes in Computer Science*, pages 142–151. Springer, Berlin, 2007.

[21] S.-H. Lai and B. C. Vemuri. Reliable and efficient computation of optical flow. *International Journal of Computer Vision*, 29(2):87–105, October 1998.

[22] C. Lei and Y. H. Yang. Optical flow estimation on coarse-to-fine region-trees using discrete optimization. In *Proc. 2009 IEEE International Conference on Computer Vision*, Kyoto, Japan, October 2009. IEEE Computer Society Press.

[23] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Seventh International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, Canada, August 1981.

[24] Y. Mileva, A. Bruhn, and J. Weickert. Illumination-robust variational optical flow with photometric invariants. In F. A. Hamprecht, C. Schnörr, and B. Jähne, editors, *Pattern Recognition*, volume 4713 of *Lecture Notes in Computer Science*, pages 152–162. Springer, Berlin, 2007.

[25] 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, March 1987.

[26] 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.

[27] H.-H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:565–593, 1986.

[28] L. Ng and V. Solo. A data-driven method for choosing smoothing parameters in optical flow problems. In *Proc. 1997 IEEE International Conference on Image Processing*, volume 3, pages 360–363, Los Alamitos, CA, USA, 1997. IEEE Computer Society.

[29] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.

[30] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[31] C. Schnörr. Segmentation of visual motion by minimizing convex non-quadratic functionals. In *Proc. Twelfth International Conference on Pattern Recognition*, volume A, pages 661–663, Jerusalem, Israel, October 1994. IEEE Computer Society Press.

[32] T. Schoenemann and D. Cremers. Near real-time motion segmentation using graph cuts. In K. Franke, K.-R. Müller, B. Nickolay, and R. Schäfer, editors, *Pattern Recognition*, volume 4174 of *Lecture Notes in Computer Science*, pages 455–464. Springer, Berlin, 2006.

[33] D. Shulman and J. Hervé. Regularization of discontinuous flow fields. In *Proc. Workshop on Visual Motion*, pages 81–86, Irvine, CA, March 1989. IEEE Computer Society Press.

[34] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *Proc. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 310–315, Maui, HI, June 1991. IEEE Computer Society Press.

[35] D. Sun, S. Roth, and M. J. Black. Secrets of optical flow estimation and their principles. In *Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, June 2010. IEEE Computer Society Press.

[36] 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.

[37] O. Tretiak and L. Pastor. Velocity estimation from image sequences with second order differential operators. In *Proc. Seventh International Conference on Pattern Recognition*, pages 16–19, Montreal, Canada, July 1984.

[38] J. van de Weijer and T. Gevers. Robust optical flow from photometric invariants. In *Proc. 2004 IEEE International Conference on Image Processing*, volume 3, pages 1835–1838, Singapore, October 2004. IEEE Signal Processing Society.

[39] A. Wedel, D. Cremers, T. Pock, and H. Bischof. Structure- and motion-adaptive regularization for high accuracy optic flow. In *Proc. 2009 IEEE International Conference on Computer Vision*, Kyoto, Japan, October 2009. IEEE Computer Society Press.

[40] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers. An improved algorithm for TV-$L^1$ optical flow computation. In D. Cremers, B. Rosenhahn, A. L. Yuille, and F. R. Schmidt, editors, *Statistical and Geometrical Approaches to Visual Motion Analysis*, volume 5604 of *Lecture Notes in Computer Science*, pages 23–45. Springer, Berlin, September 2008.

[41] J. Weickert. Theoretical foundations of anisotropic diffusion in image processing. *Computing Supplement*, 11:221–236, 1996.

[42] J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE-based computation of image motion. *International Journal of Computer Vision*, 45(3):245–264, December 2001.

[43] 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.

[44] 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.

[45] 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*, San Francisco, CA, June 2010. IEEE Computer Society Press.

[46] H. Zimmer, A. Bruhn, J. Weickert, L. Valgaerts, A. Salgado, B. Rosenhahn, and H.-P. Seidel. Complementary optic flow. In D. Cremers, Y. Boykov, A. Blake, and F. R. Schmidt, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, volume 5681 of *Lecture Notes in Computer Science*, pages 207–220. Springer, Berlin, 2009.

# A Appendix: Derivation of the Diffusion Tensor for the Complementary Regulariser

Consider the complementary regulariser from (50):

$$V(\nabla_2 u, \nabla_2 v) = \Psi_V\Big(u_{\mathbf{r}_1}^2 + v_{\mathbf{r}_1}^2\Big) + u_{\mathbf{r}_2}^2 + v_{\mathbf{r}_2}^2 \quad . \tag{57}$$

Its contributions to the Euler-Lagrange equations are given by

$$\partial_x \left(\partial_{u_x} V\right) + \partial_y \left(\partial_{u_y} V\right) \quad , \tag{58}$$

and

$$\partial_x \left(\partial_{v_x} V\right) + \partial_y \left(\partial_{v_y} V\right) \quad , \tag{59}$$

respectively. We exemplify the computation of the first expression (58). The second one then follows analogously. Let us first define the abbreviations

$$\mathbf{r}_i := (r_{i1}, r_{i2})^\top, \text{ and } \psi'_V(\mathbf{r}_i) := \Psi'_V\Big(u^2_{\mathbf{r}_i} + v^2_{\mathbf{r}_i}\Big) \,, \tag{60}$$

for $i = 1,..,2$. With their help, we compute the expressions

$$\partial_{u_x} V = 2\left(\psi'_V(\mathbf{r}_1)\, u_{\mathbf{r}_1} r_{11} + u_{\mathbf{r}_2} r_{21}\right) \,, \tag{61}$$

$$\partial_{u_y} V = 2\left(\psi'_V(\mathbf{r}_1)\, u_{\mathbf{r}_1} r_{12} + u_{\mathbf{r}_2}\, r_{22}\right) \,. \tag{62}$$

from (58). Using the fact that

$$\partial_x\left(\partial_{u_x} V\right) + \partial_y\left(\partial_{u_y} V\right) = \text{div}\,\left(\partial_{u_x} V,\, \partial_{u_y} V\right)^\top \,, \tag{63}$$

we obtain by plugging (61) and (62) into (58):

$$\partial_x\left(\partial_{u_x} V\right) + \partial_y\left(\partial_{u_y} V\right) = 2\,\text{div}\begin{pmatrix} \psi'_V(\mathbf{r}_1)\, u_{\mathbf{r}_1} r_{11} + u_{\mathbf{r}_2} r_{21} \\ \psi'_V(\mathbf{r}_1)\, u_{\mathbf{r}_1} r_{12} + u_{\mathbf{r}_2} r_{22} \end{pmatrix} \,. \tag{64}$$

By multiplying out the expressions inside the divergence expressions one ends up with

$$\partial_x\left(\partial_{u_x} V\right) + \partial_y\left(\partial_{u_y} V\right) = \tag{65}$$

$$2\,\text{div}\begin{pmatrix} \left(\psi'_V(\mathbf{r}_1)\, r^2_{11} + r^2_{21}\right) u_x + \left(\psi'_V(\mathbf{r}_1)\, r_{11} r_{12} + r_{21} r_{22}\right) u_y \\ \left(\psi'_V(\mathbf{r}_1)\, r_{11} r_{12} + r_{21} r_{22}\right) u_x + \left(\psi'_V(\mathbf{r}_1)\, r^2_{12} + r^2_{22}\right) u_y \end{pmatrix}$$

We can write above equation in diffusion tensor notation as

$$\partial_x\left(\partial_{u_x} V\right) + \partial_y\left(\partial_{u_y} V\right) = 2\,\text{div}\left(\mathbf{D}\ \nabla_2 u\right) \,, \tag{66}$$

with the diffusion tensor

$$\mathbf{D} := \begin{pmatrix} \psi'_V(\mathbf{r}_1) \cdot r^2_{11} + 1 \cdot r^2_{21} & \psi'_V(\mathbf{r}_1) \cdot r_{11} r_{12} + 1 \cdot r_{21} r_{22} \\ \psi'_V(\mathbf{r}_1) \cdot r_{11} r_{12} + 1 \cdot r_{21} r_{22} & \psi'_V(\mathbf{r}_1) \cdot r^2_{12} + 1 \cdot r^2_{22} \end{pmatrix} \,. \tag{67}$$

We multiplied the second term of each sum by a factor of 1 to clarify that the eigenvalues of $\mathbf{D}$ are $\psi'_V(\mathbf{r}_1)$ and 1, respectively. The corresponding eigenvectors are $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively, which allows to rewrite the tensor $\mathbf{D}$ as

$$\mathbf{D} = \begin{pmatrix} r_{11} & r_{21} \\ r_{12} & r_{22} \end{pmatrix}\begin{pmatrix} \psi'_V(\mathbf{r}_1)\cdot r_{11} & \psi'_V(\mathbf{r}_1)\cdot r_{12} \\ 1\cdot r_{21} & 1\cdot r_{22} \end{pmatrix} = (\mathbf{r}_1 \,|\, \mathbf{r}_2)\begin{pmatrix} \psi'_V(\mathbf{r}_1) & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \mathbf{r}_1^\top \\ \mathbf{r}_2^\top \end{pmatrix} \,. \tag{68}$$

This shows that $\mathbf{D}$ is identical to $\mathbf{D}_{\text{CR}}$ from (51), as it can be written as

$$\mathbf{D} \;=\; \psi'_V(\mathbf{r}_1)\, \mathbf{r}_1\, \mathbf{r}_1^\top + \mathbf{r}_2\, \mathbf{r}_2^\top \;=\; \Psi'_V\Big(u^2_{\mathbf{r}_1} + v^2_{\mathbf{r}_1}\Big)\, \mathbf{r}_1\, \mathbf{r}_1^\top + \mathbf{r}_2\, \mathbf{r}_2^\top \,. \tag{69}$$