# Nonlocal Regularizing Constraints in Variational Optical Flow

Joan Duran and Antoni Buades

*Department of Mathematics and Computer Science, University of Balearic Islands, Cra. de Valldemossa km. 7.5,*
*07122 Palma, Illes Balears, Spain*

Abstract:     Optical flow methods try to estimate a dense correspondence field describing the motion of the objects in an image sequence. We introduce novel nonlocal regularizing constraints for variational optical flow computation. While the use of similarity weights has been restricted to the regularization term so far, the proposed data terms permit to implicitly use the image geometry in order to regularize the flow and better locate motion discontinuities. The experimental results illustrate the superiority of the new constraints with respect to the classical brightness constancy assumption as well as to nonlocal regularization strategies.

## 1 INTRODUCTION

Since the seminal works of Horn and Schunck (Horn and Schunck, 1981) and Lucas and Kanade (Lucas and Kanade, 1981), optical flow estimation has become one of the most intensive research areas in computer vision. The main objective of optical flow methods is to compute a dense correspondence field between an arbitrary pair of images in order to capture the apparent dynamical behaviour of the objects in the scene. This is a key step in several applications like image registration, object tracking, robot navigation, stereo reconstruction, or motion based segmentation.

Generally speaking, optical flow methods can be classified into two large families. On the one hand, local techniques establish point correspondences by minimizing a distance measure matching small windows (Lowe, 2004; Yoon and Kweon, 2006). These methods commonly provide a sparse flow field due to the lack of discrimination at certain points. On the other hand, global or variational methods provide a dense solution through the minimization of an energy in which the regularization term interconnects all the pixels of the image, thus filling-in the flow wherever no sufficient information is available (Brox et al., 2004; Zach et al., 2007; Zimmer et al., 2009).

Variational optical flow methods require a constraint imposing the preservation of certain image features over the trajectories. In this regard, the classical hypothesis is the *brightness constancy assumption*, which was already considered in the Horn-Schunck and Lucas-Kanade models. However, this constraint is very limiting because is not invariant to illumination changes. Robustness can be recovered by using photometric invariant constraints, such as the *gradient constancy assumption* (Brox et al., 2004), higher-order derivatives (Papenberg et al., 2006), patch-based measures (Vogel et al., 2013), or alternative color spaces (Zimmer et al., 2009; Mileva et al., 2007). Several works (Alvarez et al., 2002; Kennedy and Taylor, 2015) have also handled the characterization of occlusions, which is one of the major challenges in realistic scenarios.

Traditionally, variational techniques include a linearization of the warped image in the data term in order to minimize the energy. This linearization is only valid for small displacements and, as a consequence, the optimization is embedded in a coarse-to-fine warping scheme (Black and Anandan, 1996; Mémin and Pérez, 1998) to better capture large motions. On the contrary, the linearization is postponed to the numerical scheme in several works (Nagel and Enkelmann, 1986; Brox et al., 2004; Brox and Malik, 2011), illustrating significative performance gain. Other approaches completely depart from coarse-to-fine strategies. For instance, Steinbrücker et al. (Steinbrücker et al., 2009) decouple the data and the regularization terms by a quadratic relaxation and the optimization problem is directly solved at the finest scale by alternating two global minimizations while decreasing the decoupling parameter.

Computing the displacement field is in general an ill-posed problem and, thus, *a priori* knowledge on the sought solution is required. This prior usually

takes the form of spatial smoothness which promotes smoothing in regions of coherent motion while permitting flow discontinuities. Nonlocal strategies have also been proposed as regularization terms (Werlberger et al., 2010; Ranftl et al., 2014).

While the use of nonlocal similarity has been restricted to the regularization term, we propose two new fidelity terms for optical flow estimation that make use of such a similarity configuration. The proposed terms permit to implicitly use the image geometry in order to regularize the flow and better locate motion discontinuities. The first term regularizes the flow by extending the classical brightness constancy assumption to similar pixels. That is, the flow for a certain pixel should be able to correctly match the color of its most similar pixels. The second term aims at replacing the classical assumption and no longer matches points along the trajectory. It uses a weight family across the two images in order to transfer patch similarity to the flow. This might be seen as an alternative for combining classical optical flow and block matching techniques.

The rest of the paper is organized as follows. In Section 2, we review the state of the art in variational optical flow computation. We derive the two novel nonlocal regularizing constraints in Section 3. Section 4 details the optimization strategy used for computing the flow field. In Section 5, we display some experimental results and we finish with some conclusions in Section 6.

## 2 STATE OF THE ART

Let $I : \Omega \times [0,T] \to \mathbb{R}$ be an image sequence, where $\Omega$ is a rectangular domain in $\mathbb{R}^2$ and $I(\mathbf{x},t)$ denotes the intensity value at pixel $\mathbf{x} = (x_1, x_2) \in \Omega$ and time $t \in [0,T]$. Let us also denote the two-dimensional displacement field by $\mathbf{u} : \Omega \times [0,T] \to \mathbb{R}^2$, where $\mathbf{u}(\mathbf{x},t) = (u_1(\mathbf{x},t), u_2(\mathbf{x},t))^\top$.

Variational optical flow methods include a data term $E_d(\mathbf{u})$, which measures the deviation from some prescribed constraints, and a regularization term $E_r(\mathbf{u})$, which quantifies the smoothness of the flow field. Therefore, the solution is computed as the minimizer of the energy functional

$$E(\mathbf{u}) = E_d(\mathbf{u}) + \lambda E_r(\mathbf{u}), \qquad (1)$$

where $\lambda$ is a trade-off parameter that balances the contribution of each term to the whole energy. The main differences among the variational models proposed in the literature rely on the choice of $E_d$ and $E_r$, and on the numerical strategies used for solving the resulting optimization problem.

### 2.1 Data-Fidelity Terms

#### 2.1.1 The Brigthness Constancy Assumption

The most widely used data-fidelity term is the brightness constancy assumption according to which the intensity of each pixel remains constant throughout the motion, i.e.,

$$I(\mathbf{x} + \mathbf{u}(\mathbf{x},t), t+1) - I(\mathbf{x},t) = 0, \quad \forall \mathbf{x} \in \Omega. \quad (2)$$

The main difficulty in the above constraint is due to the nonlinearity in the term $I(\mathbf{x} + \mathbf{u}(\mathbf{x},t), t+1)$, which involves complex computational stages. In order to tackle this issue, equation (2) is locally linearized by a first-order Taylor expansion, yielding the so-called *optical flow constraint*:

$$\nabla I(\mathbf{x},t) \cdot \mathbf{u}(\mathbf{x},t) + I_t(\mathbf{x},t) = 0, \quad \forall \mathbf{x} \in \Omega, \quad (3)$$

with $\nabla I = \left( \frac{\partial I}{\partial x_1}, \frac{\partial I}{\partial x_2} \right)$ and $I_t = \frac{\partial I}{\partial t}$. From now on, we drop the dependency of all variables over $t$, and we further write $I_1 = I(\cdot, t+1)$ and $I_0 = I(\cdot, t)$. It is worth noticing that equation (3) is only valid for small displacements or very smooth images. The standard technique to cope with large displacements is to embed the minimization in a coarse-to-fine warping (Black and Anandan, 1996; Mémin and Pérez, 1998). Several other approaches (Nagel and Enkelmann, 1986; Brox et al., 2004; Brox and Malik, 2011) use the nonlinear formulation in (2), which holds for motions of arbitrary magnitude, and postpone any linearization to the numerical scheme. Importantly, Brox et al. (Brox et al., 2004) showed that the solution at each level of the coarse-to-fine strategy used with (3) can be interpreted as a fixed point in the optimization of (2), thus both formulations are essentially equivalent in practice.

In their seminal work, Horn and Schunck (Horn and Schunck, 1981) used a quadratic function to penalize the residuals in the optical flow constraint. However, it is well known that the $L^2$ norm is not robust to outliers and occlusions. Several authors addressed this issue by replacing the quadratic error function with a robust formulation in either the nonlinear or the linearized brightness constancy assumption. The most widely used alternatives are the $L^1$ norm (Zach et al., 2007), the Charbonnier function $\varphi(s^2) = \sqrt{s^2 + \varepsilon^2}$ (Brox et al., 2004), which is a differentiable approximation of the $L^1$ norm, and the nonconvex Lorentzian function $\varphi(s^2) = \log(1 + s^2/2\sigma^2)$ (Black and Anandan, 1996).

#### 2.1.2 The Gradient Constancy Assumption

The classical brightness constancy assumption fails when additive illumination changes occur in the

scene. In order to overcome such a drawback, Brox et al. (Brox et al., 2004) introduced the so-called gradient constancy assumption in the variational framework. Instead of imposing constancy to the image brightness, the authors assumed that the image gradients remain constant under the displacement, i.e.,

$$\nabla I_1(\mathbf{x} + \mathbf{u}(\mathbf{x})) - \nabla I_0(\mathbf{x}) = 0, \quad \forall \mathbf{x} \in \Omega. \quad (4)$$

Interestingly, Wedel et al. (Wedel et al., 2009) obtained similar invariance properties by imposing brightness constancy on the textured components of the image sequence.

Despite the gain in robustness with respect to additive illumination changes, gradient constancy poses additional shortcomings: it is much more sensitive to noise than brightness, it performs poorly in smooth regions, and it does not handle local scale changes or rotations as the brightness constancy does. The constraints (2) and (4) are commonly combined in the data term to take advantage of their complementary invariance properties, leading to better flow estimations than if one of both is solely imposed (Brox et al., 2004; Brox and Malik, 2011).

More complicated features than the gradient have been used so far. In this setting, Papenberg et al. (Papenberg et al., 2006) investigated constancy conditions for higher-order features like the Laplacian or the Hessian. In the end, the experimental results illustrated that the gradient constraint (4) introduces the required illumination invariance without being as sensitive to noise as higher-order structures.

### 2.1.3 Window Regularized Constraints

Patch-based data terms make the optical flow more robust, especially to noise, since better characterize the image structure. In their seminal work, Lucas and Kanade (Lucas and Kanade, 1981) already integrated local information into the optical flow constraint (3) through a Gaussian filtering. Based on this, several authors (Bruhn et al., 2005) assumed that the displacement is almost constant over a neighbourhood around each point, i.e.,

$$\int_\Omega K_\rho(\mathbf{x} - \mathbf{y}) \, \psi \left( |I_1(\mathbf{y} + \mathbf{u}(\mathbf{x})) - I_0(\mathbf{y})| \right) d\mathbf{y}, \quad \forall \mathbf{x} \in \Omega,$$
$$(5)$$

where $K$ is a convolution kernel of size $\rho$. Note that (5) regularizes the classical brightness constancy assumption isotropically. Although this filtered-data constraint can be advantageous for very noisy sequences, it significantly blurs motion discontinuities, where this assumption fails (Zimmer et al., 2011).

## 2.2 Regularization Terms

Computing the displacement field from the previously described data terms is in general an ill-posed problem since there is no enough information to recover the optical flow at each point in the domain. Some *a priori* knowledge on the sought solution is thus required. This prior usually takes the form of spatial smoothness. In this case, the regularization term should be designed in such a way that promotes smoothing in regions of coherent motion while preserves flow discontinuities at the boundaries of moving objects. The trade-off between both scopes is in practice addressed by $\lambda$ in (1). In several works (Zimmer et al., 2011), the spatial regularization is extended to the temporal axis by assuming smoothness across consecutive image frames.

A broad class of regularizers penalizes first-order differences of the vector field through the energy

$$\int_\Omega \phi(\nabla u_1, \nabla u_2) \, d\mathbf{x}. \quad (6)$$

In the Horn-Schunck model, the $L^2$ norm, i.e. $\phi(\nabla u_1, \nabla u_2) = |\nabla u_1|^2 + |\nabla u_2|^2$, is used. However, it is well known that the square function oversmoothes the discontinuities of the flow. A large variety of robust penalty functions, such as subquadratic penalties (Black and Anandan, 1996; Mémin and Pérez, 1998), has been proposed instead. One of the most popular choices is the total variation (TV) (Rudin et al., 1992), a regularization technique that allows discontinuities yet it disfavours the solution to have oscillations. Either the classical TV (Zach et al., 2007; Wedel et al., 2009) defined as $\phi(\nabla u_1, \nabla u_2) = |\nabla u_1| + |\nabla u_2|$ or its differentiable variant (Brox et al., 2004; Brox and Malik, 2011) given by $\phi(\nabla u_1, \nabla u_2) = \sqrt{|\nabla u_1|^2 + |\nabla u_2|^2 + \varepsilon^2}$, where $\varepsilon > 0$ is a small constant that avoids the non-differentiability at zero, are widely used. However, the most relevant shortcoming of TV is the *staircasing effect*, i.e., the tendency to produce flat regions separated by artificial edges. These annoying artifacts can be almost avoided by using, for instance, the Huber norm (Werlberger et al., 2009).

An important improvement of the Horn-Schunck model was achieved by Nagel et al. (Nagel and Enkelmann, 1986), who introduced anisotropic, intensity-driven regularization penalizing oscillations in the flow field according to the direction of the intensity gradients of the image. On this basis, several methods (Alvarez et al., 2000; Werlberger et al., 2009) use anisotropic regularization terms in the form of

$$\int_\Omega g(\nabla I_0) \, \phi(\nabla u_1, \nabla u_2) \, d\mathbf{x}, \quad (7)$$

where $g$ is a spatially varying, decreasing function defined in terms of the gradient of the image. Thus, the regularization is reduced at image edges, since one assumes that there is a greater likelihood of a flow discontinuity there, and promoted inside smooth image regions. Zimmer et al. (Zimmer et al., 2009) weighted the direction of smoothing in terms of the data constraint rather than the image gradient.

The first-order priors arising from (6) and (7) often introduce a bias towards piecewise constant motions in textured areas. In order to tackle this issue, several authors (Trobin et al., 2008; Vogel et al., 2013) proposed second-order flow regularizations which favor piecewise affine vector fields. In addition, nonlocal strategies have been recently investigated (Werlberger et al., 2010; Ranftl et al., 2014). This class of methods uses the coherence of neighboring pixels to enforce similar motion patterns, yielding

$$\int_{\Omega} \int_{\mathcal{N}(\mathbf{x})} \omega(\mathbf{x}, \mathbf{y}) \phi(\mathbf{u}(\mathbf{y}) - \mathbf{u}(\mathbf{x})) \, d\mathbf{y} \, d\mathbf{x}. \qquad (8a)$$

The support weights $\omega(\mathbf{x}, \mathbf{y})$ are commonly based on spatial closeness and intensity similarity as follows:

$$\omega(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{h_s^2} - \frac{\|I_0(\mathbf{x}) - I_0(\mathbf{y})\|^2}{h_c^2}\right), \qquad (8b)$$

where $h_s$ and $h_c$ are filtering parameters that measure how fast the weights decay with increasing spatial distance or dissimilarity of colors, respectively.

## 3 TWO NEW OPTICAL FLOW CONSTRAINTS

In this section, we derive two new data constraints that permit using implicitly the image geometry in order to regularize the flow and better locate flow discontinuities. Motion patterns are enforced by means of the coherence of similar pixels. The resemblance between points is evaluated by comparing a whole window around each pixel, which is more reliable than the single pixel comparison (Buades et al., 2005).

### 3.1 Nonlocal Brightness Constancy Assumption

The first term regularizes the nonlinear brightness constancy assumption (2) by introducing nonlocal similarity. We reasonably assume that if two pixels in the source image are very similar, then the displacement assigned to one of them should work reasonably well for the other one. We thus propose the following

energy term:

$$\int_{\Omega} \int_{\Omega} \omega(\mathbf{x}, \mathbf{y}, I_0(\mathbf{x}), I_0(\mathbf{y})) \cdot$$
$$\cdot \psi(|I_1(\mathbf{y} + \mathbf{u}(\mathbf{x})) - I_0(\mathbf{y})|) \, d\mathbf{y} \, d\mathbf{x}. \qquad (9)$$

The support weights in (9) measure the similarity between patches centered at $\mathbf{x}$ and $\mathbf{y}$ in $I_0$ as follows:

$$\omega(\mathbf{x}, \mathbf{y}, I_0(\mathbf{x}), I_0(\mathbf{y})) = \frac{1}{\Gamma(\mathbf{x})} \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{h_s^2}\right) \cdot$$
$$\cdot \exp\left(-\frac{d_\rho(I_0(\mathbf{x}), I_0(\mathbf{y}))}{h_c^2}\right), \qquad (10a)$$

where $d_\rho$ denotes the distance between patches, i.e.,

$$d_\rho(I_0(\mathbf{x}), I_0(\mathbf{y})) = \int_{\Omega} G_\rho(\mathbf{z}) |I_0(\mathbf{x} + \mathbf{z}) - I_0(\mathbf{y} + \mathbf{z})|^2 d\mathbf{z}, \qquad (10b)$$

and $\Gamma(\mathbf{x})$ is a normalization factor given by

$$\Gamma(\mathbf{x}) = \int_{\Omega} \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{h_s^2} - \frac{d_\rho(I_0(\mathbf{x}), I_0(\mathbf{y}))}{h_c^2}\right) d\mathbf{y}. \qquad (10c)$$

In this framework, $G_\rho$ is a Gaussian kernel such that weights are significant only if a Gaussian window around $\mathbf{y}$ looks like the corresponding Gaussian window around $\mathbf{x}$. Furthermore, $h_s$ and $h_c$ act as filtering parameters controlling the decay of the weights as a function of the spatial and intensity patch-based distance, respectively. In the end, the average made between very similar regions preserves the integrity of the image but reduces its small fluctuations, which contain noise. Note that the weights defined in (10) satisfy the usual conditions $0 < \omega(\mathbf{x}, \mathbf{y}, I_0(\mathbf{x}), I_0(\mathbf{y})) \leq 1$ and $\int_{\Omega} \omega(\mathbf{x}, \mathbf{y}, I_0(\mathbf{x}), I_0(\mathbf{y})) \, d\mathbf{y} = 1$ for any $\mathbf{x} \in \Omega$, but the normalization factor breaks down their symmetry between two given points in the domain.

It is worth noticing that, by defining the weights as $\omega(\mathbf{x}, \mathbf{x}, I_0(\mathbf{x}), I_0(\mathbf{x})) = 1$ and $\omega(\mathbf{x}, \mathbf{y}, I_0(\mathbf{x}), I_0(\mathbf{y})) = 0$ for any $\mathbf{y} \neq \mathbf{x}$, one recovers the classic brightness constancy assumption. Moreover, compared to the isotropic formulation (5), the proposed adaptive regularization avoids the blurring of the flow close to motion discontinuities while regularizing it.

Figure 1 graphically illustrates the constraint proposed in (9). We demand two neighbouring pixels having a similar window configuration to have a similar flow. This is a softer assumption than the one by the nonlocal regularization (8), which imposes image details into the final flow (see Figure 6 in Section 5).

The constraint in (9) might be problematic for pixels with similar color having different motion. However, the weight function in (10) contains a spatial weighting imposing such a condition only for spatially closed pixels, and not sharing only the same
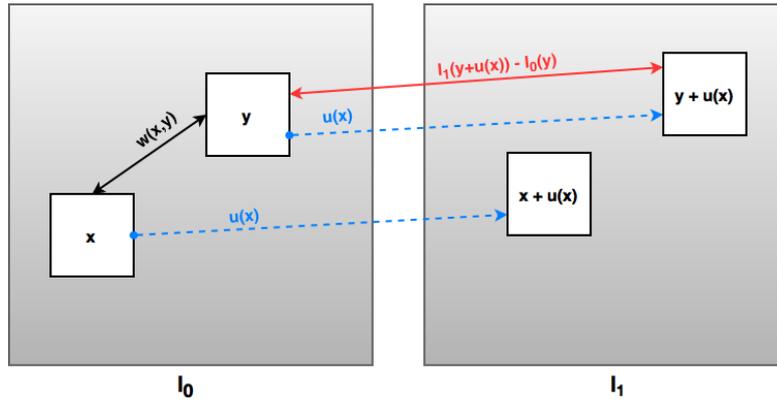
Figure 1: Graphical explanation of the Nonlocal Brightness Constancy Assumption (9). If a small window centered at $\mathbf{x}$ is similar to a small window centered at $\mathbf{y}$ in the sense described by the support weights (10), then the displacement of $\mathbf{x}$ should work for $\mathbf{y}$. As a consequence, the intensity values $I_1(\mathbf{y} + \mathbf{u}(\mathbf{x}))$ and $I_0(\mathbf{y})$ should correspond to each other.

pixel color but the color of an entire window around. Both aspects of weight definition makes (9) robust to the existence of pixels with similar intensity values but different displacements.

Finally, let us mention that the nonlocal brightness constancy term is closely related to a bilateral correction window and is in fact inspired by Yoon et al. (Yoon and Kweon, 2006) and Xiao et al. (Xiao et al., 2006), who used similar ideas for local matching in stereo.

## 3.2 Nonlocal Matching Assumption

The second new-proposed term aims at replacing the classic brightness constancy assumption. It uses a weight family across the two images of the pair in order to transfer window similarity to the displacement field as follows:

$$\int_\Omega \int_\Omega \omega(I_0(\mathbf{x}), I_1(\mathbf{y})) \psi(|I_1(\mathbf{x} + \mathbf{u}(\mathbf{x})) - I_1(\mathbf{y})|) \, d\mathbf{y} \, d\mathbf{x}. \tag{11}$$

Let us emphasize that the brightness constancy assumption (2) cannot be obtained from (11) with any weight distribution. Actually, the proposed term no longer imposes a constraint on the motion trajectories but a nonlocal self-similarity.

The weights in (11) measure the similarity between a patch centered at $\mathbf{x}$ in $I_0$ and another one centered at $\mathbf{y}$ in $I_1$ as follows:

$$\omega(I_0(\mathbf{x}), I_1(\mathbf{y})) = \frac{1}{\Gamma(\mathbf{x})} \exp\left(-\frac{d_\rho(I_0(\mathbf{x}), I_1(\mathbf{y}))}{h_c^2}\right), \tag{12a}$$

where the distance between patches is computed as

$$d_\rho(I_0(\mathbf{x}), I_1(\mathbf{y})) = \int_\Omega G_\rho(\mathbf{z}) |I_0(\mathbf{x} + \mathbf{z}) - I_1(\mathbf{y} + \mathbf{z})|^2 d\mathbf{z} \tag{12b}$$

and the normalization factor is

$$\Gamma(\mathbf{x}) = \int_\Omega \exp\left(-\frac{d_\rho(I_0(\mathbf{x}), I_1(\mathbf{y}))}{h_c^2}\right) d\mathbf{y}. \tag{12c}$$

Note that the difference between the weights defined in (10) and those given in (12) is that the latter only depend on the color similarity since the spatial closeness is not considered. This is because we do not want the constraint (11) to be limited to small displacements but to deal with pixels being relatively far from each other, being the only limitation the linearization of the numerical scheme if applicable.

Figure 2 illustrates the constraint defined in (11). If pixels $\mathbf{x}$ in $I_0$ and $\mathbf{y}$ in $I_1$ are similar, then we can expect the pixel in $I_1$ assigned to $\mathbf{x}$ by the flow field to be similar to $\mathbf{y}$. This is not a hard constraint since we do not demand the pixel $\mathbf{x}$ to be matched with $\mathbf{y}$, but only to share a similar color.

The nonlocal matching assumption wants to introduce patch comparison or block matching into the optical flow variational formulation. Local block matching methods compute motion by matching a small window around any pixel in the first image with the window in the second image minimizing a certain cost. These methods fail when the window is not distinctive enough to be matched. This might be due to the lack of texture inside the window but also to the presence of several copies of the window in the second image. In that case, which is known as the *aperture problem*, local methods are not able to decide among the different candidates. The use of the weight configuration between patches of both images in (12) permits introducing block matching as a soft constraint into the energy. Therefore, (11) helps in suppressing artifacts due to wrong motion estimations and noise.
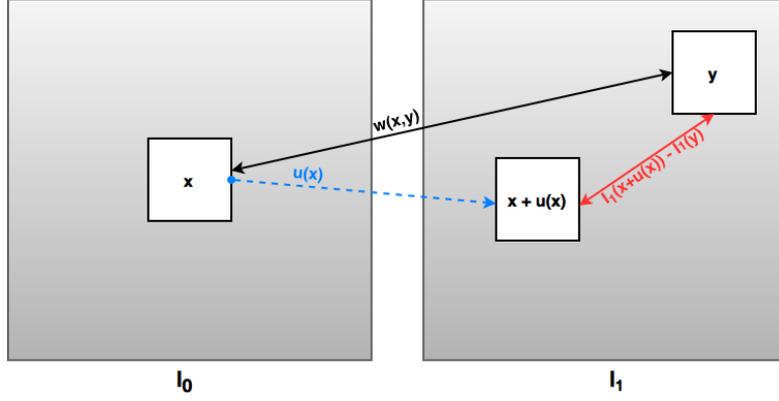
Figure 2: Graphical illustration of the Nonlocal Matching Assumption in (11). If a small window centered at $\mathbf{x}$ in $I_0$ is similar to a small window centered at $\mathbf{y}$ in $I_1$, as described by the support weights (12), then the displacement field should assign to $\mathbf{x}$ a pixel in $I_1$ with an intensity value similar to $\mathbf{y}$. As a consequence, the values $I_1(\mathbf{x} + \mathbf{u}(\mathbf{x}))$ and $I_1(\mathbf{y})$ should be similar.

## 3.3 The Energy Functionals

We specify now the full energy in the form (1). Throughout this work, we employ quadratic functions, i.e. $\psi(s) = s^2$, to penalize the deviations from the prescribed assumptions in (9) and (11). Even though knowing the shortcomings of this choice, which obviously may affect the quality of the results, our aim is to compare the validity of the novel constraints with respect to the classical brightness constancy assumption rather than pursuing the best penalization. The latter will be part of future work.

Let us notice that the proposed constraints are nonlinear in $\mathbf{u}$ because of the warpings $I_1(\mathbf{y} + \mathbf{u}(\mathbf{x}))$ and $I_1(\mathbf{x} + \mathbf{u}(\mathbf{x}))$. We linearize both expressions using Taylor expansions as follows:

$$
\begin{aligned}
g(\mathbf{u}) := \ & I_1(\mathbf{y} + \mathbf{u}^0(\mathbf{x})) - I_0(\mathbf{y}) \\
& + \langle \nabla I_1(\mathbf{y} + \mathbf{u}^0(\mathbf{x})), \mathbf{u}(\mathbf{x}) - \mathbf{u}^0(\mathbf{x}) \rangle, \\
f(\mathbf{u}) := \ & I_1(\mathbf{x} + \mathbf{u}^0(\mathbf{x})) - I_1(\mathbf{y}) \\
& + \langle \nabla I_1(\mathbf{x} + \mathbf{u}^0(\mathbf{x})), \mathbf{u}(\mathbf{x}) - \mathbf{u}^0(\mathbf{x}) \rangle,
\end{aligned}
\tag{13}
$$

with $\mathbf{u}^0$ being a close approximation to $\mathbf{u}$.

In addition to the data-fidelity terms, a critical part in a variational model is the prior. The regularization is responsible for the propagation of the flow from boundaries to homogenous regions. This propagation relies on the spatial coherence of natural images, thus pixels belonging to the same object are supposed to have almost the same flow. We incorporate the total variation (Rudin et al., 1992) as regularization for the two novel constraints.

Using (13), the final linearized nonlocal brightness constancy energy is

$$
\begin{aligned}
E_\gamma^l(\mathbf{u}) := \ & \sum_{m=1}^{2} \int_\Omega |\nabla u_m(\mathbf{x})| \, d\mathbf{x} \\
& + \frac{\gamma}{2} \int_\Omega \int_\Omega \omega(\mathbf{x}, \mathbf{y}, I_0(\mathbf{x}), I_0(\mathbf{y})) \, (g(\mathbf{u}))^2 \, d\mathbf{y} \, d\mathbf{x},
\end{aligned}
\tag{14a}
$$

while the linearized nonlocal matching energy is

$$
\begin{aligned}
E_\delta^l(\mathbf{u}) := \ & \sum_{m=1}^{2} \int_\Omega |\nabla u_m(\mathbf{x})| \, d\mathbf{x} \\
& + \frac{\delta}{2} \int_\Omega \int_\Omega \omega(I_0(\mathbf{x}), I_1(\mathbf{y})) \, (f(\mathbf{u}))^2 \, d\mathbf{y} \, d\mathbf{x}.
\end{aligned}
\tag{14b}
$$

# 4 NUMERICAL MINIMIZATION

## 4.1 Convex Relaxation of the Energies

Inspired by Zach et al. (Zach et al., 2007), we relax the minimization of (14) by introducing an auxiliary variable $\mathbf{v}$ which decouples the data and regularization terms as

$$
\begin{aligned}
& \sum_{m=1}^{2} \int_\Omega |\nabla u_m(\mathbf{x})| \, d\mathbf{x} + \frac{1}{2\theta} \|\mathbf{u} - \mathbf{v}\|^2 \\
& + \frac{\gamma}{2} \int_\Omega \int_\Omega \omega(\mathbf{x}, \mathbf{y}, I_0(\mathbf{x}), I_0(\mathbf{y})) (g(\mathbf{v}))^2 d\mathbf{y} \, d\mathbf{x}
\end{aligned}
\tag{15a}
$$

and

$$
\begin{aligned}
& \sum_{m=1}^{2} \int_\Omega |\nabla u_m(\mathbf{x})| \, d\mathbf{x} + \frac{1}{2\theta} \|\mathbf{u} - \mathbf{v}\|^2 \\
& + \frac{\delta}{2} \int_\Omega \int_\Omega \omega(I_0(\mathbf{x}), I_1(\mathbf{y})) (f(\mathbf{v}))^2 d\mathbf{y} \, d\mathbf{x}.
\end{aligned}
\tag{15b}
$$

Therefore, one can compute the solution by alternating minimizations:

i) Fixed $\mathbf{v}$, solving (15) with respect to $\mathbf{u}$ is a TV-based problem, the solution of which is computed using Chambolle's projection algorithm (Chambolle, 2004).

ii) Fixed $\mathbf{u}$, the minimizer of (15) with respect to $\mathbf{v}$ can be computed explicitly due to the quadratic penalization in the data-fidelity energy terms.

## 4.2 Computation of the Weights

For computational purposes, the nonlocal interaction is limited to pixels at a certain distance, the so-called *research window*. More precisely, given a parameter $\nu > 0$, we redefine the weights (10) and (12) as

$$\omega(\mathbf{x},\mathbf{y},I_0(\mathbf{x}),I_0(\mathbf{y})) = \frac{1}{\Gamma(\mathbf{x})} \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{h_s^2}\right) \cdot$$
$$\cdot \exp\left(-\frac{1}{h_c^2}\sum_{\mathbf{z}\in\mathcal{N}_0}|I_0(\mathbf{x}+\mathbf{z})-I_0(\mathbf{y}+\mathbf{z})|^2\right) \tag{16}$$

and

$$\omega(I_0(\mathbf{x}),I_1(\mathbf{y})) = \frac{1}{\Gamma(\mathbf{x})} \cdot$$
$$\cdot \exp\left(-\frac{1}{h_c^2}\sum_{\mathbf{z}\in\mathcal{N}_0}|I_0(\mathbf{x}+\mathbf{z})-I_1(\mathbf{y}+\mathbf{z})|^2\right) \tag{17}$$

if $\|\mathbf{x}-\mathbf{y}\|_\infty \leq \nu$, and $\omega \equiv 0$ otherwise. Here, $\mathcal{N}_0$ is a rectangular window centered at $\mathbf{0}$, the so-called *comparison window*. The Gaussian kernel $G_\rho$ introduced in (10) and (12) is not considered in practice as it is only necessary when the size of the windows increase considerably. After all, the weight distribution is commonly sparse.

## 4.3 Coarse-to-Fine Approach

A problem that arises with the linearizations performed in (13) is how to determine the value of $\mathbf{u}^0$ in order to allow large disparities between images. We use a coarse-to-fine scheme to reduce the distance between the objects in the scene. Furthermore, in each scale, $\mathbf{u}^0$ is iteratively refined to assure convergence.

We employ image pyramids of 5 scales with a subsampling factor of 2. The images are smoothed with a Gaussian kernel of standard deviation 1.04 before subsampling. Beginning with the coarsest level, we solve (15) at each scale of the pyramid and propagate the solution to the next one as $\mathbf{u}^{s-1}(\mathbf{x}) = 2\mathbf{u}^s(0.5\mathbf{x})$.

Every intermediate solution is used as the initialization in the following scale. In each scale, we introduce 5 intermediate steps to update $\mathbf{u}^0$ and warp $I_1$. At the beginning of a new scale, $\mathbf{v}$ is initialized with $\mathbf{u}$ and, at the coarsest one, $\mathbf{u}$ starts with $\mathbf{0}$.

The displacement to be detected must be small at the coarsest scale. In this respect, one drawback of the pyramidal approach is that the method cannot estimate the motion of small objects undergoing large displacements, since these may disappear in the coarsest scales. However, let us emphasize that this is not a limitation of the new-proposed constraints, but it is a matter of the linearization and the optimization strategy we have chosen.

Spatial image and flow derivatives are discretized using central differences and forward differences, respectively, with Neumann boundary conditions. Furthermore, we use bicubic interpolation to warp the image $I_1$ and its derivatives. At each warp, the minimization procedure alternates one step of the fixed-point scheme to update $\mathbf{u}$ (Chambolle, 2004) with the explicit computation of $\mathbf{v}$. As stopping criterion we use a tolerance value of $10^{-6}$ for the relative error between two consecutive iterations. Anyway, we stop the algorithm after 1000 iterations even if the tolerance is not reached.

The optical flow is computed on the grayscale images and the sizes of the research and comparison windows used in the support weights are fixed to $21 \times 21$ and $7 \times 7$ pixels, respectively. Furthermore, we integrate a median filter of size $7 \times 7$ into the numerical scheme to increase the robustness to sampling artifacts in the image data (Wedel et al., 2009). For all models under comparison, this filtering step is applied after each warp.

Finally, let us mention that the computational costs of the two new-proposed methods given in (15) are equivalent to the computational cost of the classical TV-L2 method, that is, the one penalizing the linearized counterpart of the brightness constancy assumption with the Euclidean norm. Indeed, we have only to perform an extra weight computation at the beginning of each scale of the multi-resolution pyramid. Since this might be easily parallelized, the increase in running time is negligible.

## 5 EXPERIMENTAL RESULTS

This section aims at comparing the two novel data terms with the classic brightness constancy assumption. We evaluate the methods with the Middlebury benchmark (Barker et al., 2011) with known ground truth, so that we can determine the optimal param-

| Reference | Classical Constraint (3) | Nonlocal Constraint (9) |
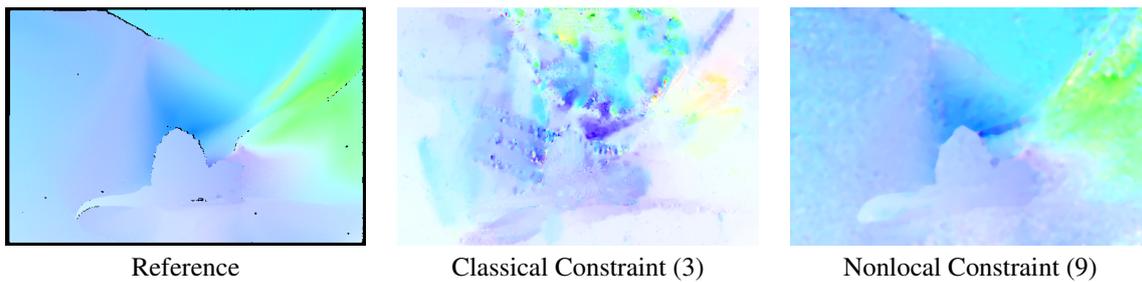
Figure 3: Optical flow estimation without regularization. For the classic linearized constraint (3) we added a TV regularization term limiting its role as much as possible. While (3) leads to an unstable flow, the nonlocal brightness constancy assumption (9) permits the computation of an admissible solution without any prior.



| Reference | Classic energy | $E_\gamma^l$ | $E_\delta^l$ |
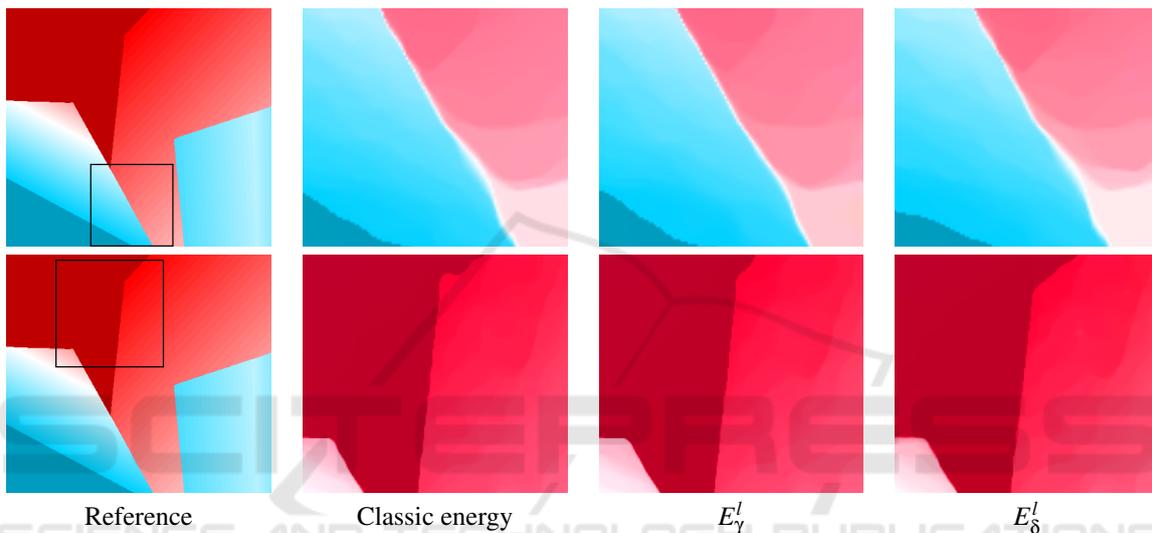
Figure 4: Reference and obtained flow fields for the "Venus" sequence. The AEPE are 0.313, 0.309, and 0.310 for the classic, $E_\gamma^l$, and $E_\delta^l$ energies, respectively.

eters in terms of the lowest average end-point error (AEPE). The color scheme used in all experiments to represent the orientation and magnitude of the optical flow is the same as that of Barker et al. (Barker et al., 2011). In order to fairly compare all variational models, we use a numerical scheme similar to (15) for the model based on the brightness constancy assumption (2), but penalizing its linearized counterpart, given in (3), with the $L^2$ norm.

It is well known that it is not possible to determine the flow from (3) since the number of parameters to be estimated is larger than the number of linearly independent equations. However, the linearized nonlocal brightness constancy constraint (9) allows us to get an explicit solution for the motion without any prior. Figure 3 displays the obtained flows in both cases – since (3) is ill-posed, we added the TV regularization but reducing its role as much as possible. We observe that, as expected, the classical constraint leads to an unstable flow as the regularization vanishes. The constraint (9) is able to give an admissible result without

any prior instead.

Figures 4 and 5 provide the flow fields obtained for the "Venus" and "Rubberwhale" sequences, respectively. We also display the corresponding AEPE values in order to numerically compare the results with the ground truth. We have excluded from this measure the pixels in the occlusions, which are available for the Middlebury benchmark. The AEPE values were computed for the whole image and not only for the close-ups showed in the figures. While there is hardly any visual difference between the flow fields estimated by the classic energy and $E_\delta^l$ in the first row of Figure 4, the latter is convincingly better in the close-ups from the second row. Indeed, $E_\delta^l$ is able to correctly estimate the flow at the boundaries of the objects – see the slope at the top of these images. On the other hand, the nonlocal brightness constancy assumption in $E_\gamma^l$ identifies the gap in the middle of the Venus image better than the others, as highlighted in the close-ups from the first row. Similar behavior of $E_\gamma^l$ can be observed in the Rubberwhale sequence. In

|  |  |  |  |
|---|---|---|---|
| Reference | Classic energy | $E_\gamma^l$ | $E_\delta^l$ |
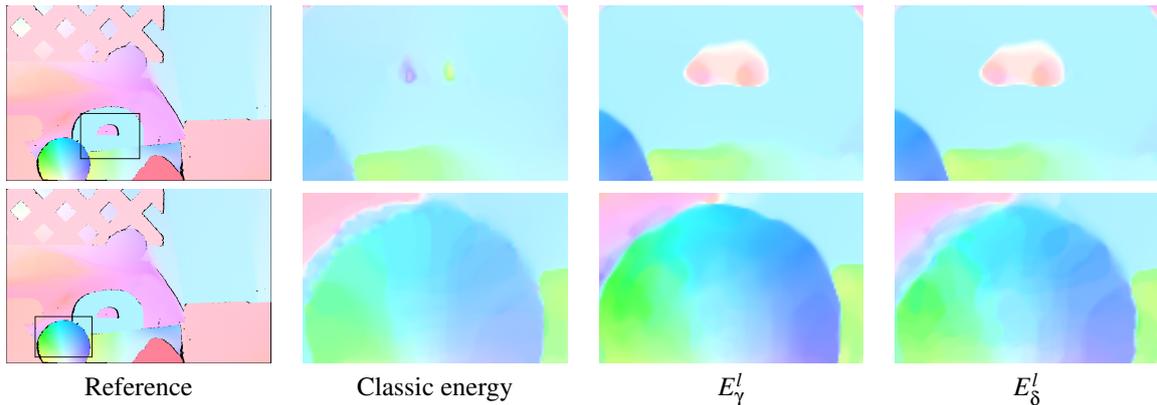
Figure 5: Reference and obtained flow fields for the "Rubberwhale" sequence. The AEPE are 0.209, 0.154, and 0.199 for the classic, $E_\gamma^l$, and $E_\delta^l$ energies, respectively.



|  |  |  |  |  |
|---|---|---|---|---|
| Frame | Reference | NLTV | $E_\gamma^l$ | $E_\delta^l$ |

Figure 6: Source frame, reference and obtained flow fields for the "Venus" sequence. The classic optical flow constraint (3) is regularized using nonlocal strategies (NLTV). Observe that NLTV copies the geometry and texture of the source frame into the estimated flow, which does not happen with the nonlocal constraints $E_\gamma^l$ and $E_\delta^l$.

this case, both $E_\gamma^l$ and $E_\delta^l$ detect the hole in the letter 'e', which is not the case with the classic constraint. The results in the second row of Figure 5 show that $E_\gamma^l$ provides the best visual and numerical performance. In the end, the two nonlocal regularizing data terms show a significantly discriminative potential when compared with the classic energy. Furthermore $E_\delta^l$ is proved to replace the brightness constancy assumption efficiently without matching points along the trajectory.

We finally compare the two new-proposed nonlocal constraints with the nonlocal regularization (8) used as prior jointly with the optical flow constraint given in (3) (Werlberger et al., 2010; Ranftl et al., 2014). Figure 6 displays close-ups of the estimated motion for the Venus sequence. We observe that the nonlocal regularization forces the geometry and texture of the image into the flow field, identifying wrong motion patterns. On the contrary, the proposed data terms use the image geometry correctly to regularize the flow and better locate flow discontinuities.

## 6 CONCLUSIONS

In this paper, we have introduced two nonlocal constraints for optical flow estimation. The image geometry is used to propose regularized data-fidelity terms making the flow computation more robust and able to better locate motion discontinuities. The experimental results have illustrated their superiority with respect to the classic brightness constancy assumption. The results also demonstrate that image self-similarity can be better taken advantage of in the data-fidelity terms than in the regularization prior. For the moment, we limited ourselves to illustrate the performance of each term separately, the combination of them will be object of future research.

The limitations of this work are in the optimization strategy rather than in the models themselves. We have linearized the constraints, forcing us to embed the optimization in a coarse-to-fine warping. Future work will mainly concentrate on postponing the linearization to the numerical scheme and using the nonlinear formulations directly, which will require a careful minimization strategy.

## ACKNOWLEDGEMENTS

## REFERENCES

Alvarez, L., Deriche, R., Papadopoulo, T., and Sánchez, J. (2002). Symmetrical dense optical flow estimation with occlusions detection. In *Proc. European Conf. Computer Vision (ECCV)*, volume 2350 of *Lecture Notes in Comp. Sci.*, pages 721–735, Copenhagen, Denmark.

Alvarez, L., Weickert, J., and Sánchez, J. (2000). Reliable estimation of dense optical flow fields with large displacements. *Int. J. Comput. Vis.*, 39(1):41–56.

Barker, S., Scharstein, D., Lewis, J., Roth, S., Black, M., and Szeliski, R. (2011). A database and evaluaton methodology for optical flow. *Int. J. Comput. Vis.*, 92(1):1–31.

Black, M. and Anandan, P. (1996). The robust estimation of multiple motions: Parametric and piecewise smooth flow fields. *Comput. Vis. Image Underst.*, 63(1):75–104.

Brox, T., Bruhn, A., Papenberg, N., and Weickert, J. (2004). High accuracy optical flow estimation based on a theory for warping. In *Proc. European Conf. Computer Vision (ECCV)*, volume 3024 of *Lecture Notes in Comp. Sci.*, pages 25–36, Prague, Czech Republic.

Brox, T. and Malik, J. (2011). Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(3):500–513.

Bruhn, A., Weickert, J., and Schnörr, C. (2005). Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *Int. J. Comput. Vis.*, 61(3):211–231.

Buades, A., Coll, B., and Morel, J.-M. (2005). A review of image denoising algorithms, with a new one. *SIAM Multiscale Model. Simul.*, 4(2):490–530.

Chambolle, A. (2004). An algorithm for total variation minimization and applications. *J Math. Imaging Vis.*, 20(1–2):89–97.

Horn, B. and Schunck, B. (1981). Determining optical flow. In *Technical Symposium East*, pages 319–331. International Society for Optics and Photonics.

Kennedy, R. and Taylor, C. (2015). Optical flow with geometric occlusion estimation and fusion of multiple frames. In *Proc. Int. Conf. Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, volume 8932 of *Lecture Notes in Comp. Sci.*, pages 364–377, Hong Kong.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.*, 60(2):91–110.

Lucas, B. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proc. Int. Joint Conf. Artificial Intelligence*, pages 674–679, Vancouver, BC, Canada.

Mémin, E. and Pérez, P. (1998). Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Trans. Image Process.*, 7(5):703–719.

Mileva, Y., Bruhn, A., and Weickert, J. (2007). Illumination-robust variational optical flow with photometric invariants. In *Proc. DAGM Symposium on Pattern Recognition*, volume 4713 of *Lecture Notes in Comp. Sci.*, pages 152–162, Heidelberg, Germany.

Nagel, H.-H. and Enkelmann, W. (1986). An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(5):565–593.

Papenberg, N., Bruhn, A., Brox, T., Didas, S., and Weickert, J. (2006). Highly accurate optic flow computation with theoretically justified warping. *Int. J. Comput. Vis.*, 67(2):141–158.

Ranftl, R., Bredies, K., and Pock, T. (2014). Non-local total generalized variation for optical flow estimation. In *Proc. European Conf. Computer Vision (ECCV)*, volume 8698 of *Lecture Notes in Comp. Sci.*, pages 439–454, Zurich, Switzerland. Springer Berlin Heidelberg.

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

Steinbrücker, F., Pock, T., and Cremers, D. (2009). Large displacement optical flow computation without warping. In *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, pages 1609–1614, Kyoto, Japan.

Trobin, W., Pock, T., Cremers, D., and Bischof, H. (2008). An unbiased second-order prior for high-accuracy motion estimation. In *Proc. DAGM Symposium on Pattern Recognition*, volume 5096 of *Lecture Notes in Comp. Sci.*, pages 396–405, Munich, Germany.

Vogel, C., Roth, S., and Schindler, K. (2013). An evaluation of data costs for optical flow. In *Proc. German Conference Pattern Recognition (GCPR)*, volume 8142 of *Lecture Notes in Comp. Sci.*, pages 343–353, Saarbrücken, Germany.

Wedel, A., Pock, T., Zach, C., Cremers, D., and Bischof, H. (2009). An improved algorithm for TV-L1 optical flow. In *Proc. Statistical and Geometrical Approached to Visual Motion Analysis*, volume 5604 of *Lecture Notes in Comp. Sci.*, pages 23–45, Dagstuhl Castle, Germany.

Werlberger, M., Pock, T., and Bischof, H. (2010). Motion estimation with non-local total variation regularization. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 2464–2471, San Francisco, CA, USA.

Werlberger, M., Trobin, W., Pock, T., Bischof, H., Wedel, A., and Cremers, D. (2009). Anisotropic Huber-L1 optical flow. In *Proc. British Machine Vision Conference (BMVC)*, pages 108.1–108.11, London, UK.

Xiao, J., Cheng, H., Sawhney, H., Rao, C., and Isnardi, M. (2006). Bilateral filtering-based optical flow es-

timation with occlusion detection. In *Proc. European Conf. Computer Vision (ECCV)*, volume 3951 of *Lecture Notes in Comp. Sci.*, pages 211–224, Graz, Austria.

Yoon, K. and Kweon, I. (2006). Adaptive support-weight approach for correspondece search. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(4):650–656.

Zach, C., Pock, T., and Bischof, H. (2007). A duality based approach for realtime TV-L1 optical flow. In *Proc. DAGM Symposium on Pattern Recognition*, volume 4713 of *Lecture Notes in Comp. Sci.*, pages 214–223, Heidelberg, Germany.

Zimmer, H., Bruhn, A., and Weickert, J. (2011). Optic flow in harmony. *Int. J. Comput. Vis.*, 93(3):368–388.

Zimmer, H., Bruhn, A., Weickert, J., Valgaerts, L., Salgado, A., Rosenhahn, B., and Seidel, H.-P. (2009). Complementary optic flow. In *Proc. Int. Conf. Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, volume 5681 of *Lecture Notes in Comp. Sci.*, pages 207–220, Bonn, Germany.