Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

Optical flow estimation on image sequences with differently exposed frames

Tomas Bengtsson
Tomas McKelvey
Konstantin Lindström

SPIE.

Optical flow estimation on image sequences with differently exposed frames

Tomas Bengtsson,[a,*] Tomas McKelvey,[a] and Konstantin Lindström[b]
[a]Chalmers University of Technology, Department of Signals and Systems, Göteborg 412 96, Sweden
[b]Volvo Car Corporation, Göteborg 405 31, Sweden

Abstract. Optical flow (OF) methods are used to estimate dense motion information between consecutive frames in image sequences. In addition to the specific OF estimation method itself, the quality of the input image sequence is of crucial importance to the quality of the resulting flow estimates. For instance, lack of texture in image frames caused by saturation of the camera sensor during exposure can significantly deteriorate the performance. An approach to avoid this negative effect is to use different camera settings when capturing the individual frames. We provide a framework for OF estimation on such sequences that contain differently exposed frames. Information from multiple frames are combined into a total cost functional such that the lack of an active data term for saturated image areas is avoided. Experimental results demonstrate that using alternate camera settings to capture the full dynamic range of an underlying scene can clearly improve the quality of flow estimates. When saturation of image data is significant, the proposed methods show superior performance in terms of lower endpoint errors of the flow vectors compared to a set of baseline methods. Furthermore, we provide some qualitative examples of how and when our method should be used. [DOI: 10.1117/1.OE.54.9.093103]

Keywords: optical flow estimation; temporal regularization; multiple camera settings; high-dynamic range.

Paper 150343 received Mar. 17, 2015; accepted for publication Jul. 28, 2015; published online Sep. 10, 2015.

1 Introduction

Optical flow (OF) estimation is the task of finding the pixel-wise two-dimensional (2-D) planar flow between two consecutive frames in an image sequence.[1–3] It provides dense, nonrigid motion information which is crucial for several applications. For driver assistance systems in vehicles, OF and related methods provide useful low-level information to help track the surrounding traffic scenario.[4] For segmentation of objects, such as vehicles and pedestrians, using the estimated flow data instead of the image intensity data itself has shown superior results.[5] Estimated flow fields are also commonly used for image registration directly, for example, to register time-sequences of medical imaging data.[6] In a large survey of camera-based pedestrian detection systems, lack of contrasts in the captured image data is identified as a core reason for lack of quality in the object motion estimates.[7] Two issues, related to the camera sensor setup, contribute to the undesired lack of contrasts between objects in certain image regions. First, the dynamic range of the scene may be higher than that of the camera sensor, which causes saturation of the sensor in certain image regions of a given frame.[8,9] Second, different objects in the scene may fundamentally be poorly contrasted relative to each other in the sensed spectral band. The negative effects of both issues can be mitigated by alternating the exposure setting of the camera system between successive frames. This paper presents a novel framework to estimate flow data for sequences of differently exposed images. We make use of recent advances to temporal regularization of OF and utilize more than two images for each flow field estimate. The framework allows any number of input images to be used. For the experiments presented here, we use four images with two different exposure settings that are used alternately every other frame. Based on this setup, intended to address the issue of saturation in the image data, a set of flow estimation methods are proposed. In addition to the proposed methods, a set of baseline methods that only use one exposure setting for all the frames are discussed and included in the experimental evaluation. Before we make connections to related work and summarize the contributions of the paper, some background on OF methods is presented.

1.1 *Optical Flow Foundations*

The success of OF estimation in general depends upon a set of factors. A brief overview of the historical development of OF methods helps to highlight these. For the moment, consider all images to be equally exposed, which is the conventional case. For each point in the reference image, the objective is to estimate the corresponding flow vector which describes the motion of the point to its new location at a later time instance. This OF estimation problem is formulated as finding the minimizer of a given cost expression. The modeling of the cost expression is often based on the so called brightness constancy assumption (BCA), which states that the brightness intensity of any given point in an image is unchanged at its later locations, or equivalently, along its motion trajectory. Thus, a data penalty term is formulated on the image intensity data to relate the points of two images by a 2-D flow field that should be estimated. A good flow field estimate naturally corresponds to a low data cost. As long as deviations from the BCA are small between

*Address all correspondence to: Tomas Bengtsson, E-mail: tomas.bengtsson@chalmers.se

consecutive frames, this approach can result in robust flow estimates.[10] For complex illumination scenarios in which the BCA is invalid, the temporal illumination variations either need to be included in the model explicitly (in the simplest example as a spatially invariant offset) or the data cost must be based on transformed image data to an image function that is robust to these variations.[11,12]

Since there are two unknown components (horizontal and vertical) of the flow vector for each pixel, it is clear that a pixelwise data cost in itself is insufficient to formulate a problem with a unique solution if the pixel data is scalar-valued. Furthermore, even multivalued pixel intensities, for instance, three channel color images where the BCA is applied to each channel individually, do not provide a stable estimation problem due to the fact that certain image regions, such as the interiors of homogenous objects, lack texture in all channels simultaneously. There are two primary methods to regularize the flow estimation problem and thus force the existence of a unique solution. Both apply a condition on the spatial distribution of the flow to enforce the solution to be spatially (piecewise) smooth. This is based on the fact that pixels on a specific object exhibit a similar flow as long as the shape of the object is not deformed. One approach is to formulate the spatial smoothness condition locally using patches of pixels,[13] whereas the other more common choice is to globally formulate the condition.[1,4] Mixtures of local and global formulations are also possible.[14] In the current, state-of-the-art variational mathematical approaches to OF estimation, the primary choice is to use a global regularization term as part of a total cost functional to be minimized.[3,15] That choice is adopted in our work.

For image sequences with relatively simple motion, i.e., without motion of a large magnitude or complex motion patterns, and with minor illumination changes between the images, traditional OF methods that consist of a pointwise data term and a global spatial regularization term in combination with robust penalty functions provide competitive results.[10] This is the case for, e.g., the Middlebury benchmark,[16] whose dataset only contains motion of small magnitudes as well as minor illumination variations due to the controlled experimental setup. The recent advancement of OF benchmark datasets has gone hand in hand with sophisticated developments of OF methods. For instance, re-evaluating existing OF methods that ranked high on the classic Middlebury benchmark on the more challenging MPI Sintel benchmark dataset[17] showed that certain methods that performed well on the former benchmark do not necessarily perform well on the latter benchmark, highlighting certain drawbacks of those methods. (The benchmark rankings are available in Refs. 18 and 19.) MPI Sintel includes several sequences with large and complex motions of small-scale image structures, as well as challenging natural illumination variations, all of which have been missing in the Middlebury dataset. When large motion is present, a coarse-to-fine multiresolution strategy is generally adopted to avoid convergence to local minima, as well as to reduce the computational complexity.[20,21] The flow solution at a coarse image resolution is used to initialize the estimation problem at the next, finer scale in what is typically called a warping scheme. However, if the magnitude of the motion is larger than the size of the corresponding image structure, multiresolution strategies fail.[22] Error-propagation across resolution scales occurs, resulting in failure to resolve areas of fine details that tend to be over-smoothed by the regularization. Therefore, recent works have taken into account local image features in a novel fashion, which have led to improved estimation results. Deviations from sparse, prematched features are penalized in an additional term of the total cost functional.[15,23–26] For large motions, this term helps to steer the flow estimate out of local optima. For complex and small-scale motions, the extra feature matching term helps by reducing the relative influence of the spatial smoothness term, which does not hold for those regions. As an alternative to multiresolution techniques altogether, sparse-to-dense estimation techniques have shown strong performance and in particular, they avoid error-propagation across scales.[27,28]

1.2 *Related Work*

Just as the flow solution can be accurately assumed a priori to be piecewise smooth, many image sequences contain scenarios that exhibit temporally smooth flow. Attempts to exploit this, by use of combined spatial and temporal regularizations, have shown improved performance compared to methods employing only spatial regularization.[29–31] Recently, a novel method for temporal regularization along motion trajectories was introduced.[31] The key insight implemented in their paper is that temporal comparison of the flow should not be made to the flow at the same pixel location in the next frame, but rather at the new location to which the given pixel has moved. To this end, a parametrization of the flow components as flow increments relative to the pixel locations of a reference frame was introduced. This concept is adopted in the methods proposed here.

To address the issue of natural illumination variations between consecutive images, a number of alternative data terms that use transformed image data have been proposed. A simple alternative is to formulate the data term based on the assumption of constant image intensity gradients rather than on the BCA.[3] Another approach is to use structure-texture decomposition and supply texture-enhanced images or pure texture-images as inputs.[32] As a final example, the use of census transformed images has been proposed to obtain an illumination-robust data term.[11,33] There are also promising methods that explicitly model the illumination variations and argue that it is, in fact, undesired to discard brightness and contrast magnitudes which, e.g., the census transform does.[12] However, none of the above address the issue of saturation in the image data. In this context, there is some recent work on using alternate exposure images for OF estimation. Sellent et al.[34] use two short-exposed images and an intermediate long-exposed image and use information such as the direction of the motion blur to enhance the flow estimation performance. However, no dynamic range aspects are treated in their work. Hafner et al.[35] on the contrary, present a method that utilizes a sequence of images of different exposure durations which jointly estimates a high-dynamic range image (HDR) as well as the OF. They are the first authors to publish a method which provides dense flow field estimates within the context of HDR image reconstruction,[8,9] and show that the joint approach benefits both the HDR image and flow estimates. The estimation of the HDR image along with the flow field allows image-driven, anisotropic spatial regularization to be used in the flow estimation steps, however,

it adds computational cost if the primary interest is motion estimation.

1.3 *Contributions*

In this work, the task is to estimate OF between a specific reference frame and the next frame which is exposed using a different camera setting. The use of alternating exposure durations, commonly used, e.g., to capture the full dynamic range in HDR scenarios, is one example of adjustable camera settings. Another example, suitable for some applications, is to use flash illumination every other frame. An advantage of using flash instead of a long exposure duration in order to capture low-intensity data is that the issue of motion blur can be mitigated. On the other hand, for alternating exposure durations, the mutually nonsaturated image regions can be photometrically aligned simply by scaling the sensor data with the inverse of the exposure duration, which can be useful for the flow estimation. As an illustration of the negative effects of saturation in the image data, consider the example shown in Fig. 1. Due to clipping of the low-intensity data in the input images, the resulting flow estimate fails to capture the full extent of the moving person. Specifically, the flow estimate is poor in the saturated lower image region.

We propose a formulation of the OF data term which operates on pairs of frames that have been equally exposed, but still provides flow estimates between consecutive image pairs, thanks to enforcing temporal smoothness across the incremental flow terms. All of the data from the image sequence is thus merged into one estimation task to avoid undesired situations where the effect of the regularization terms dominate the flow solution. In summary, the contributions of the paper are:

- A solution method to the extended OF estimation problem that allows image sequences with differently exposed frames to be used.
- Quantitative evaluation of flow estimates obtained from a set of proposed OF data terms, including comparisons to conventional methods, on altered data from public datasets.
- Qualitative examples on the performance degradation of flow estimates caused by saturated image data.

1.4 *Outline of the Paper*

Different generative data models and their implications are discussed in Sec. 2. The variational formulation of the OF estimation problem is then presented in Sec. 3. A set of baseline methods that use a single exposure setting for all frames are presented in Sec. 4 and the proposed methods that use image sequences with differently exposed frames are described in Sec. 5. Experimental results are given in Sec. 6 and Sec. 7 concludes the paper with a presentation of directions for further research.

2 Generative Data Models

The formulation of the OF estimation problem should ideally depend on how the input data, i.e., the image sequence, is generated. To begin with, formulating a data cost term (as part of the total cost functional to be minimized) between two images taken with the same camera setting is straightforward and corresponds to the conventional OF case. However, whether or not it is possible to include a data term between image pairs taken with different camera settings depends on if there is a reasonable model to relate these image pairs. Consider that an image, $\tilde{I}_f$, is generated according to the camera model

$$\tilde{I}_f(\mathbf{x}) = \mathrm{CRF}\{\Phi_f[R(\mathbf{x}) + N_f(\mathbf{x})]\}, \tag{1}$$

where $\mathbf{x} = (x, y) \in \Omega \subset \mathbb{R}^2$ is the image domain, and $R(\mathbf{x})$ is the (filtered) illuminance incident on the sensor for the specific lightning condition of the imaged scene at the time instance of the image $\tilde{I}_f$. The noise term $N_f(\mathbf{x})$ can, e.g., represent the quantization noise, and the function Φ_f models the effect of the specific camera settings used to generate the image. Finally, the camera response function (CRF) is a pointwise function whose argument is the raw sensor data. The CRF clips the light exposure $X_f \triangleq \Phi_f[R(\mathbf{x}) + N_f(\mathbf{x})]$ outside of the operating interval of the sensor, limited by its dynamic range.[36] It also typically encodes the data using a concave function, approximately a gamma power law with a typical value of $\gamma \approx 1/2.2$,[37,38] e.g., for image storage. For such a case, $\mathrm{CRF}(X_f) = [c(X_f)]^\gamma$, where c is a clipping function. In an HDR scenario, the dynamic range of X_f is higher than that of the camera sensor.

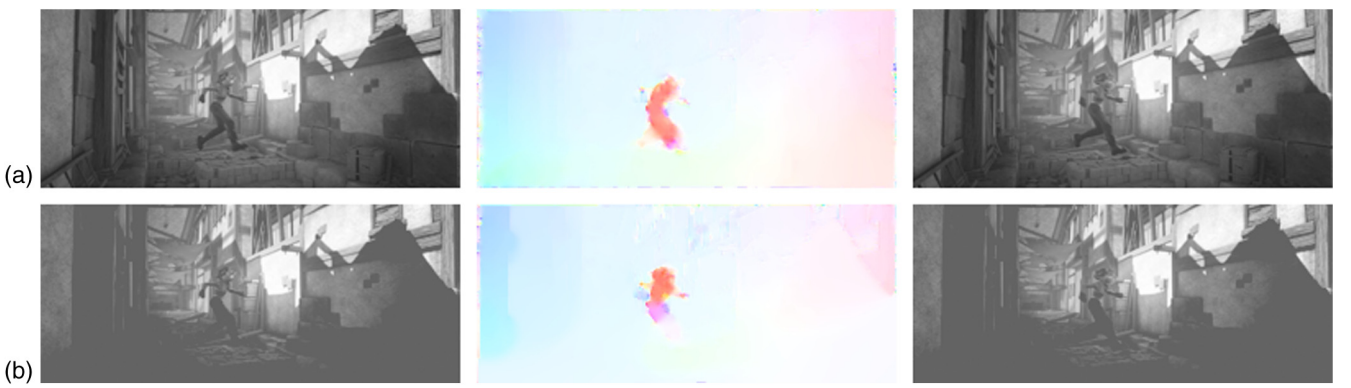

Fig. 1 Effect of saturated regions in the image data on the resulting flow estimate. (a) A pair of consecutive frames from MPI Sintel[17] and the flow estimate from a method of this paper. (b) The same two frames with their low-intensity data clipped under a fixed threshold value and the flow estimate based on these frames. Due to the lack of contrasts that arises in the lower regions of the frames, the flow estimates fail to capture the full extent of the person that moves in the sequence.

Under the BCA and considering nonoccluded regions, another image, $\tilde{I}_{f+1}$, of the same scene can be related to $\tilde{I}_f$ through

$$\tilde{I}_{f+1}[\mathbf{x}+\mathbf{w}_f(\mathbf{x})] = \mathrm{CRF}\{\Phi_{f+1}[R(\mathbf{x})+N_{f+1}(\mathbf{x})]\}, \tag{2}$$

where $\mathbf{w}_f(\mathbf{x})$ denotes the displacement, or flow, of point $\mathbf{x}$ in $\tilde{I}_f$. If the images $\tilde{I}_f$, $\tilde{I}_{f+1}$ are generated using the same camera settings, then $\Phi_{f+1} = \Phi_f$, which enables us to formulate a data term between $\tilde{I}_{f+1}$ and $\tilde{I}_f$ in order to estimate the OF field $\mathbf{w}_f$ that relates those images. Next, consider an HDR case where the full dynamic range is captured by the use of different exposure durations. A long exposure duration is used to capture low-intensity data which, however, also results in overexposed, saturated high-intensity regions. Similarly, a short exposure duration is used to capture high-intensity data, which leads to clipping of underexposed low-intensity data. Generally speaking, if the image pair $\tilde{I}_f$, $\tilde{I}_{f+1}$ are taken with different exposure durations δt_1 and δt_2 and all other camera settings are equal, then they can be related through the imaged scene, characterized by $R(\mathbf{x})$, according to the model

$$\begin{aligned}\tilde{I}_f(\mathbf{x}) &= \mathrm{CRF}[\delta t_1 R(\mathbf{x}) + N_f(\mathbf{x})],\\ \tilde{I}_{f+1}[\mathbf{x}+\mathbf{w}_f(\mathbf{x})] &= \mathrm{CRF}[\delta t_2 R(\mathbf{x}) + N_{f+1}(\mathbf{x})].\end{aligned} \tag{3}$$

The nonsaturated data can be photometrically aligned by inverting the effect of the CRF and scaling with the inverse of the respective exposure durations. Thus, this pair of differently exposed images can be used to formulate a data term between the mutually nonsaturated image regions. On the contrary, if the alternative camera setting is such that there exists no model to relate the nonsaturated data in Eqs. (1) and (2), photometric alignment is not possible. For instance, if $\tilde{I}_{f+1}$ is captured with a flash but $\tilde{I}_f$ is not, the flash will illuminate the scene in a spatially varying manner, causing changes to $R(\mathbf{x})$ that are difficult to model. It may still be possible to implicitly align the images photometrically by using transformed image functions, e.g., using the census transform.[11,33] Such an approach, however, is outside the scope of this paper.

In the remainder of the paper, the notation without tilde, I_f, is used for frames in image sequences. For image sequences with differently exposed frames, the use of this notation implies that the given frames are photometrically aligned for any sequence where there exists a mathematical model to relate them.

3 Variational Optical Flow Estimation

In variational OF methods, input images are seen as time-samples of a continuous image intensity function $I(\mathbf{x},t)$. A given frame, $I_f = I(\mathbf{x},t_f)$, can be related to the next frame at a later time instance, $I_{f+1} = I(\mathbf{x},t_{f+1})$, e.g., under the BCA. Without loss of generality, assume that $t_{f+1} = t_f + 1$. Then the BCA can be stated mathematically as

$$I_{f+1}[\mathbf{x}+\mathbf{w}_f(\mathbf{x})] - I_f(\mathbf{x}) = 0, \tag{4}$$

where $\mathbf{w}_f = \mathbf{w}(\mathbf{x},t_f) = [u(\mathbf{x},t_f), v(\mathbf{x},t_f)]$ is a flow field containing the horizontal and vertical unknown flow functions at time t_f. In particular, it describes the integrated flow from the time instance of a given frame to the next. Deviations from the BCA are small for nonoccluded areas with approximately constant illumination properties at times t_f and t_{f+1}. When this holds, a good estimate of the flow should likewise keep the magnitude of the left hand side of Eq. (4) small. Thus, the equality of Eq. (4) is relaxed and the left hand side is taken as the OF data cost.

A general form of the total cost functional that should be minimized for our variational OF estimation is

$$\begin{aligned}E(\{\mathbf{w}_f\}) &= E_D + \alpha_S E_S + \alpha_T E_T =\\ &= \int_\Omega (F_D + \alpha_S F_S + \alpha_T F_T)\mathrm{d}\mathbf{x},\end{aligned} \tag{5}$$

where the data term is denoted by E_D, the spatial and temporal regularization terms on the flow are E_S and E_T with respective weights $\alpha_S, \alpha_T > 0$, and $\{\mathbf{w}_f\}$ is a set of flow variables. The corresponding pointwise terms, F_D, F_S, and F_T, are introduced to simplify the notation, and they are also central in the derivation of the Euler–Lagrange (E–L) equations that provide the necessary conditions for the flow fields $\{\mathbf{w}_f\}$ that minimize Eq. (5).[39] Traditionally, OF is often estimated without any temporal condition on the flow.[1,2] Given a sequence of image frames $\{I_f\} = I(\mathbf{x}, t = \{t_f\})$, estimating the flow between two consecutive frames then only involves those particular frames in the data term. For example, consider estimating the flow that relates I_2 and I_3, denoted $\mathbf{w}_2 = (u_2, v_2)$. For nonoccluded image regions, under the condition of brightness constancy, $I_3(\mathbf{x}+\mathbf{w}_2) = I_2(\mathbf{x})$. Thus, the data term

$$F_D = \Psi[|I_3(\mathbf{x}+\mathbf{w}_2) - I_2(\mathbf{x})|^2], \tag{6}$$

is formulated in such a way as to minimize the pointwise differences under some penalty function Ψ over the image domain. In this work, $\Psi(z^2) = (z^2+\epsilon^2)^{1/2}$, $\epsilon = 10^{-3}$ is adopted to measure the data term deviations using a convex, differentiable L^1-norm approximation which is robust to modeling errors, e.g., due to occluded objects in the images.[3,4,40,41]

To set the stage for OF estimation on sequences with differently exposed frames, which is the ultimate goal here, temporal regularization is now introduced to the same task as the current example, i.e., to estimate $\mathbf{w}_2$. A total of $\mathcal{F} = 4$ frames are treated at once. In addition to I_2, I_3 themselves, I_1 and I_4 are used because they are closest in time. When dealing with $\mathcal{F} > 2$ frames, the most direct way to relate them pairwise is according to

$$\begin{aligned}&\Psi\{[I_2(\mathbf{x}+\tilde{\mathbf{w}}_1) - I_1(\mathbf{x})]^2\},\\ &\Psi\{[I_3(\mathbf{x}+\tilde{\mathbf{w}}_2) - I_2(\mathbf{x})]^2\},\\ &\Psi\{[I_4(\mathbf{x}+\tilde{\mathbf{w}}_3) - I_3(\mathbf{x})]^2\}.\end{aligned} \tag{7}$$

However, although $\tilde{\mathbf{w}}_2 = \mathbf{w}_2$ is the desired flow, previous results that are confirmed in our work show that this is a poor parametrization on components.[31] This is because penalizing differences between $\tilde{\mathbf{w}}_1$, $\tilde{\mathbf{w}}_2$, and $\tilde{\mathbf{w}}_3$ correspond to comparing flow at the same spatial location from one time instance to the next, which does not make sense. What is desired, rather, is to compare the flow of an object, e.g., $\tilde{\mathbf{w}}_2(\mathbf{x})$, to the flow of the same object at its new location in the next frame, i.e., $\tilde{\mathbf{w}}_2(\mathbf{x}) \approx \tilde{\mathbf{w}}_3[\mathbf{x}+\tilde{\mathbf{w}}_2(\mathbf{x})]$ and similarly

$\tilde{\mathbf{w}}_2(\mathbf{x}) \approx \tilde{\mathbf{w}}_1[\mathbf{x} + \tilde{\mathbf{w}}_1^*(\mathbf{x})]$ should hold, where $\tilde{\mathbf{w}}_1^*$ denotes backward flow.[30] To this end, an alternative, more suitable parametrization is achieved by expressing the flow fields as increments relative to specific locations in a reference frame $I_{f_{\text{ref}}}$, here $f_{\text{ref}} = 2$.[31] Then the terms of the data cost become

$$
\begin{aligned}
F_{D12} &= \Psi\{[I_2(\mathbf{x}) - I_1(\mathbf{x} - \mathbf{w}_1)]^2\},\\
F_{D23} &= \Psi\{[I_3(\mathbf{x} + \mathbf{w}_2) - I_2(\mathbf{x})]^2\},\\
F_{D34} &= \Psi\{[I_4(\mathbf{x} + \mathbf{w}_2 + \mathbf{w}_3) - I_3(\mathbf{x} + \mathbf{w}_2)]^2\}.
\end{aligned} \tag{8}
$$

In the implementation of the numerical minimization that follows, this parametrization ensures that all flow increments are indexed relative to the reference pixel grid $\mathbf{x}$ of I_2. Noninteger arguments of I_1, I_3, I_4 that result from subpixel precision OF estimates are evaluated using bicubic interpolation. The total data cost is

$$
F_D = \theta_{12} F_{D12} + \theta_{23} F_{D23} + \theta_{34} F_{D34}, \tag{9}
$$

where $\theta_{12}(\mathbf{x})$, $\theta_{23}(\mathbf{x})$, $\theta_{34}(\mathbf{x}) \geq 0$ are the scalar weight functions that enable to weigh well exposed and poorly exposed image regions differently. In particular, they are used here to remove the influence of a given data cost term in image regions where one image contains saturated data, but could in addition be set according to the noise properties of a specified camera setup with a related generative data model. With the parametrization in Eq. (8), the temporal cost term enforced along flow trajectories simply becomes

$$
F_T = \Psi(\|\mathbf{w}_2 - \mathbf{w}_1\|^2) + \Psi(\|\mathbf{w}_3 - \mathbf{w}_2\|^2), \tag{10}
$$

where $\|\cdot\|$ is the L^2-norm. Finally, the spatial regularization term is taken as

$$
\begin{aligned}
F_S = \Psi(&\|\nabla u_1\|^2 + \|\nabla v_1\|^2 +\\
&+ \|\nabla u_2\|^2 + \|\nabla v_2\|^2 +\\
&+ \|\nabla u_3\|^2 + \|\nabla v_3\|^2).
\end{aligned} \tag{11}
$$

The spatial term E_S sums the pointwise contributions from F_S over the spatial domain according to the L^1-approximation of Ψ as defined earlier. This means that it penalizes the (approximated) total variation, due to the gradients of the flow functions, across the image domain.[3,40–42] Compared to the L^2-norm (over the spatial domain), it is efficient at preserving flow edges. Alternative (global) spatial regularization methods are generalized total variation[11] or anisotropic methods that estimate edge orientations and aim to only smooth the flow along edges but not across.[43] Some further design choices for the spatial and temporal regularization terms are also discussed in the original work on the parametrization along flow trajectories.[31]

4 Baseline Optical Flow Methods

The focus of this work is to use alternating camera settings to overcome the negative impact saturated image data have on resulting flow estimates. Before describing the proposed methods, in this section, we consider the use of a single, fixed exposure settings for all frames in an HDR scenario, for which the dynamic range of the scene exceeds the dynamic range of the camera sensor. Thus, all input images are taken either with exposure setting I or with exposure setting II, such that one option leads to saturation in low-intensity image regions and the other leads to saturation in high-intensity image regions across the whole image sequence. To limit the scope of methods to evaluate, we exclude the plausible intermediate case which, to a lesser extent, contains saturated regions in both low and high intensity regions.

The objective here as well as for the proposed methods is to estimate $\mathbf{w}_2$, the flow at the reference frame, by minimizing a total cost functional of the form in Eq. (5). The data terms in Eqs. (6) and (9) are used to form four baseline methods:

$$
\begin{aligned}
&(\text{A})^{\text{Exp.I}}\ \theta_{23} F_{D23}, \quad \theta_{23} = 1, \quad \forall\ \mathbf{x},\\
&(\text{B})^{\text{Exp.I}}\ \theta_{12} F_{D12} + \theta_{23} F_{D23} + \theta_{34} F_{D34},\\
&\theta_{12} = \theta_{23} = \theta_{34} = 1, \quad \forall\ \mathbf{x},
\end{aligned}
$$

$$
\begin{aligned}
&(\text{A})^{\text{Exp.II}}\ \theta_{23} F_{D23}, \quad \theta_{23} = 1, \quad \forall\ \mathbf{x},\\
&(\text{B})^{\text{Exp.II}}\ \theta_{12} F_{D12} + \theta_{23} F_{D23} + \theta_{34} F_{D34},\\
&\theta_{12} = \theta_{23} = \theta_{34} = 1, \quad \forall\ \mathbf{x}.
\end{aligned}
$$

The superscripts of $(\text{A})^{\text{Exp.I}}$ and $(\text{B})^{\text{Exp.I}}$ state that the input frames are all taken with exposure setting I, and analogously for the superscripts of $(\text{A})^{\text{Exp.II}}$ and $(\text{B})^{\text{Exp.II}}$. The methods $(\text{B})^{\text{Exp.I}}$ and $(\text{B})^{\text{Exp.II}}$ contain the temporal and spatial regularization terms Eqs. (10) and (11) in their total cost functionals. The methods $(\text{A})^{\text{Exp.I}}$ and $(\text{A})^{\text{Exp.II}}$ contain only one flow term, thus there is no temporal regularization, only a spatial regularizer $F_S = \Psi(\|\nabla u_2\|^2 + \|\nabla v_2\|^2)$. Note that all the data terms have weight one over the whole image domain. By using only one camera setting, the same objects are saturated in all images which effectively removes the influence of the data term for those regions even for a nonzero weight. In fact, setting the weight to zero in saturated regions seems to give slightly worse quantitative results than not doing so, likely due to discarding the information of the boundary between the saturated and nonsaturated areas.

From here on, we consider the proposed setup where every other frame is exposed differently, i.e., the frames I_1, I_3 are captured using exposure setting I and the frames I_2, I_4 are captured using exposure setting II. In that case, directly applying unweighted versions of either the data term Eqs. (6) or (9) is out of the question. This is because pairs of images that are differently exposed are compared in these data terms. In particular, image regions that are saturated in one image but not the other would then lead to corrupted flow estimates. A sequence of differently exposed images is exemplified in Fig. 2. Four frames from the Middlebury sequence Grove2 are altered by clipping (saturating) high-intensity data in I_1, I_3 and low-intensity data in I_2, I_4, which is the same effect that would occur in an HDR scenario. To adapt the data terms in Eqs. (6) and (9) to the case of different exposure settings, what could be done is to enforce them only in areas where neither image is saturated. That is, Eq. (6) can be multiplied by a weight θ_{23} which is zero in areas where I_2 is saturated as well as in areas where I_3 is saturated, and similarly for Eq. (9). We denote the nonsaturated subsets of the image domain Ω for exposure setting I and II, respectively, by $\Omega^{\text{Exp.I}} \subset \Omega$ and $\Omega^{\text{Exp.II}} \subset \Omega$. The

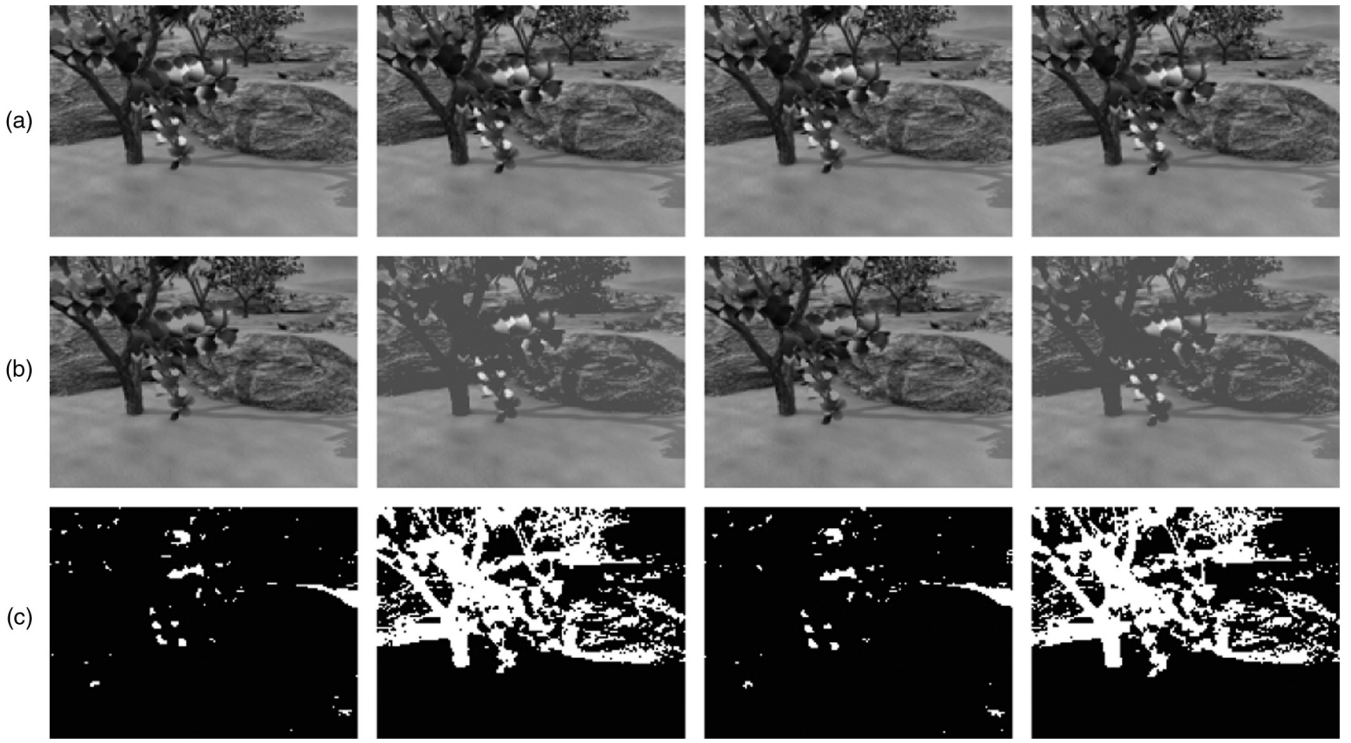

Fig. 2 The image sequence Grove2 from the Middlebury dataset, with I_1 in the leftmost column followed by I_2, I_3 and I_4 in order: (a) original images, (b) simulated high-dynamic range (HDR) sequence, used for the experimental evaluation, and (c) masks that show saturated pixels in white.

intersection $\Omega^{\text{Int}} = \Omega^{\text{Exp.I}} \cap \Omega^{\text{Exp.II}}$ is the set of points that are nonsaturated for both exposure settings. The following two data costs are given as initial methods for OF estimation on sequences with differently exposed frames:

$$\text{(A)}\ \theta_{23} F_{D23}, \quad \theta_{23} = 1, \quad \mathbf{x} \in \Omega^{\text{Int}},$$
$$\text{(B)}\ \theta_{12} F_{D12} + \theta_{23} F_{D23} + \theta_{34} F_{D34},$$
$$\theta_{12} = \theta_{23} = \theta_{34} = 1, \quad \mathbf{x} \in \Omega^{\text{Int}}.$$

We use the convention that the weights are zero for points where they are not explicitly specified. Note that, in method (B), the weighting has the unfortunate effect of discarding data correspondences in the same regions of all four frames, even though two of the frames are properly exposed and could provide valid information for the flow estimation procedure. The flow estimation performances of (A) and (B) are clearly limited by the fact that they only utilize mutually nonsaturated image regions.

5 Optical Flow Estimation on Sequences with Differently Exposed Frames

In this section, we provide a generalization of the OF data cost expression that is suitable for flow estimation on image sequences with differently exposed frames. The proposed methods are restricted to use up to four frames, indexed $f = 1, 2, 3, 4$. However, in general, there is no fixed limit to how many frames that could be included, except for the increased computational demand. As stated previously, the objective is to estimate $\mathbf{w}_2$ by minimizing a total cost functional of the form in Eq. (5). The expression for the (pointwise) data term is generalized to

$$F_D = \sum_{n=1}^{N} \theta_{p_n q_n} F_{D p_n q_n},$$
$$F_{D p_n q_n} = \Psi\{[I_{q_n}(\mathbf{x} + \mathbf{W}_{q_n}) - I_{p_n}(\mathbf{x} + \mathbf{W}_{p_n})]^2\}. \tag{12}$$

The subscripts p_n and q_n, $n = 1, \ldots, N$, describe the image pairs that are compared in the overall data term. Each data term is weighted by a function $\theta_{p_n q_n}(\mathbf{x}) \geq 0$. The upper case flow fields $\mathbf{W}_{q_n} = (U_{q_n}, V_{q_n})$ and $\mathbf{W}_{p_n} = (U_{p_n}, V_{p_n})$ are cumulative flow fields, such that for some reference frame $f_{\text{ref}} = r$,

$$\begin{cases} \mathbf{W}_f = -\mathbf{w}_f - \ldots - \mathbf{w}_{r-1}, & f < r \\ \mathbf{W}_r = 0, & \\ \mathbf{W}_f = \mathbf{w}_r + \ldots + \mathbf{w}_{f-1}, & f > r. \end{cases} \tag{13}$$

The general expression for the spatial regularizer is

$$F_S = \Psi\left(\sum_f \|\nabla u_f\|^2 + \|\nabla v_f\|^2\right), \tag{14}$$

and the temporal regularization term is

$$F_T = \sum_f \Psi(\|\mathbf{w}_{f+1} - \mathbf{w}_f\|^2), \tag{15}$$

where the summations over f include all flow increments from the first to the last frame included in the data term. The minimizer of Eq. (5), and thus the estimate of $\mathbf{w}_2$, is found iteratively by successive linearizations of the argument of Ψ in the data term Eq. (12) (i.e., the BCA), and of the nonlinear expression for Ψ.[3] At each step, the update is given by the solution to the corresponding E-L equations.[39]

The derivation of the E-L equations as well as details on the numerical implementation are given in Appendix A. What differs between all the different methods to be presented in this section is essentially the data term, specifically p_n, q_n and the weights $\theta_{p_n q_n}(\mathbf{x})$. The number of flow increment terms in use and thus the summation limits of Eqs. (14) and (15) depend on the choices of p_n, q_n. For all presented cases, $f_{\text{ref}} = 2$ is the reference frame. Furthermore, the weight functions $\theta_{p_n q_n}$ are only explicitly specified for points $\mathbf{x}$ where they are nonzero and $\theta_{p_n q_n} = 0$ holds otherwise.

When the frames I_1, I_3 are captured using exposure setting I and the frames I_2, I_4 are captured using exposure setting II, two OF cost expressions are given straightforwardly using the following data terms:

$$\text{(C)}\ \theta_{13} F_{D13}, \quad \theta_{13} = 1, \quad \forall\, \mathbf{x},$$

$$\text{(D)}\ \theta_{24} F_{D24}, \quad \theta_{24} = 1, \quad \forall\, \mathbf{x}.$$

Each of these by itself only uses the pair of similarly exposed images among the total four frames to produces a flow estimate. The data terms in (C) and (D) are both parameterized with respect to the reference frame I_2. If (C) had been taken with $f_{\text{ref}} = 1$, it would correspond to the poor but common parametrization of the sort in Eq. (7). By penalizing in F_T the difference between, e.g., $\mathbf{w}_2$ and $\mathbf{w}_3$ in the data term $F_{D24} = \Psi\{[I_4(\mathbf{x} + \mathbf{w}_2 + \mathbf{w}_3) - I_2(\mathbf{x})]^2\}$ of (D), the total flow between frames I_2 and I_4 is shared equally between the two flow terms $\mathbf{w}_2$ and $\mathbf{w}_3$. The same is also true for (C). Note that, e.g., in (D), a single flow component $\mathbf{w}_{2+3} = \mathbf{w}_2 + \mathbf{w}_3$ would be more natural and lead to a smaller dimensionality of the problem, but the expressions are given as is because then all presented cases are encompassed by the same general formulation Eq. (12) and implementation.

5.1 *Proposed Methods*

Three methods are proposed in this section. Each of these methods uses the information from all alternately exposed images $I_1, \ldots, I_4$ in some way. There should clearly be a benefit to the estimation performance if saturated image regions in I_1, I_3 contain properly exposed image data in those regions in I_2, I_4, and vice versa. To begin with, the formulation of the flow as incremental terms in the data costs for (C) and (D) allows us to directly form an additional flow estimate, the weighted sum

$$\text{(E)}\ \mathbf{w}_2^{(E)} = c(\mathbf{x})\mathbf{w}_2^{(C)} + d(\mathbf{x})\mathbf{w}_2^{(D)},$$

where $\mathbf{w}_2^{(C)}$ and $\mathbf{w}_2^{(D)}$ are the estimates of $\mathbf{w}_2$ from the methods (C) and (D), respectively. The weight terms $c(\mathbf{x})$, $d(\mathbf{x})$ are such that $c(\mathbf{x}) + d(\mathbf{x}) = 1$, $\forall\, \mathbf{x}$. For points $\mathbf{x} \notin \Omega^{\text{Exp.I}}$, $c(\mathbf{x}) = 0$, $d(\mathbf{x}) = 1$, and similarly saturated points in I_2 result in $d(\mathbf{x}) = 0$, $c(\mathbf{x}) = 1$. Otherwise, for mutually nonsaturated points $\mathbf{x} \in \Omega^{\text{Int}}$, a simply weighting is obtained by setting $c(\mathbf{x}) = d(\mathbf{x}) = 0.5$. However, for a well defined generative data model, these weight terms could additionally take into account the signal-to-noise ratio of the different exposure settings. There is a potential drawback related to the fact that (E) is formed by a sum of flow data from two separate estimation procedures. As there is no coupling between $\mathbf{w}_2$ in (C) and $\mathbf{w}_2$ in (D), the respective estimates $\mathbf{w}_2^{(C)}$ and $\mathbf{w}_2^{(D)}$ are formed using a lower frame-rate than, e.g., the estimates (A) and (B). To overcome this issue, the following data term is proposed:

$$\text{(F)}\ \theta_{13} F_{D13} + \theta_{24} F_{D24},$$

$$\theta_{13} = 1, \quad \theta_{24} = 1, \quad \mathbf{x} \in \Omega^{\text{Int}},$$

$$\theta_{13} = 2, \quad \theta_{24} = 0, \quad \{\mathbf{x} \in \Omega^{\text{Exp.I}} \text{ and } \mathbf{x} \in \Omega^{\text{Exp.II}}\},$$

$$\theta_{13} = 0, \quad \theta_{24} = 2, \quad \{\mathbf{x} \in \Omega^{\text{Exp.II}} \text{ and } \mathbf{x} \in \Omega^{\text{Exp.I}}\}.$$

Similarly to method (E), there is one data term based on the image pair I_1, I_3 taken with exposure setting I and another one for the image pair I_2, I_4 taken with exposure setting II. Here, however, the flow variable $\mathbf{w}_2$ is shared by the two data terms and method (F) is, therefore, expected to outperform (E). The choice of weights θ_{13}, θ_{24} is not obvious. Simulations show that, e.g., using weights $\theta_{13} = \theta_{24} = 1$, $\forall \mathbf{x}$, gives a slightly worse performance compared to the specified weights. Bearing in mind that the regularization weights are spatially constant, it seems advantageous to shift the weight for saturated regions in one image pair to the other image pair that contains texture in those regions. In general, $\theta_{13} = \theta_{24}$ for $\mathbf{x} \in \Omega^{\text{Int}}$ is merely a special case. Just as for c, d in (E), any choice of weights can be made based on the properties of the noise in a specified generative model for a given image sequence. If $\Omega^{\text{Exp.I}} \cup \Omega^{\text{Exp.II}} = \Omega$, the method (F) ensures that at least one of its two data terms is active in the whole image domain.

The third and final proposed method,

$$\text{(G)}\ \theta_{13} F_{D13} + \theta_{24} F_{D24} + \theta_{23} F_{D23},$$

extends method (F) by including an additional data term between mutually nonsaturated regions of the differently exposed images I_2, I_3. The weights θ_{13}, θ_{24} are set the same way as in (F) and $\theta_{23} = 1$, $\mathbf{x} \in \Omega^{\text{Int}}$. The method (G) is meaningful in cases where the nonsaturated regions can be photometrically aligned, as discussed in Sec. 2, either as preprocessing or implicitly by using a transformed image domain for the data terms. To conclude the section, an overview of the proposed methods described here, as well as of the baseline methods, is given in Table 1. The presented flow estimation methods are evaluated experimentally Sec. 6.

6 Experimental Results and Discussion

Two experimental setups are used to evaluate the presented flow estimation cost functionals. Experiment 1, which is separated into 1a and 1b due to similar experiments but on different datasets, is performed on altered data from the Middlebury[18] and MPI Sintel[19] training sets (using the "Final" render pass for the MPI Sintel sequences[17]). Both training sets consist of synthetic, animated data that is suitable for evaluation purposes since ground truth flow is available. Experiment 2 is performed on data from a prototype camera setup which includes near-infrared spectral sensitivity and where every second frame is captured with flash illumination and the remainder without. For Experiment 1, the regularization weights are set to $\alpha_T = \alpha_S/5$ for all methods that contain temporal regularization and the free parameter α_S is individually selected by a grid-search for each method,

Table 1 Summary of the baseline methods, $(A)^{Exp.I}$, $(A)^{Exp.I}$, $(B)^{Exp.II}$, $(B)^{Exp.II}$, and the proposed methods (E), (F), and (G).

Case	Description
$(A)^{Exp.I}/(A)^{Exp.II}$	The flow estimation uses two frames, I_2, I_3, both captured using exposure setting I/II.
$(B)^{Exp.I}/(B)^{Exp.II}$	The flow estimation uses four frames, $I_1, \ldots, I_4$, all captured using exposure setting I/II.
(E)	The flow estimate is given by a weighted sum of two estimates from separate methods, (C) and (D), that each uses a single pair of equally exposed frames, I_1, I_3 and I_2, I_4, respectively. For (E) and the other proposed methods, as opposed to the baseline methods, I_1, I_3 are captured using exposure setting I and I_2, I_4 are captured using exposure setting II.
(F)	Two data cost terms, based on the image pairs I_1, I_3 and I_2, I_4, respectively, are combined into one cost expression in order to achieve a coupled estimation of the desired flow field w_2.
(G)	This method extends method (F) by including an additional data cost based on the mutually nonsaturated regions of I_2, I_3.

minimizing the pixelwise endpoint error of the estimated flow $\mathbf{w}_2^{(\cdot)}(\mathbf{x})$ relative to the ground truth $\mathbf{w}_2^{gt}(\mathbf{x})$, $\mathrm{EPE} = \|\mathbf{w}_2^{(\cdot)} - \mathbf{w}_2^{gt}\|$, averaged over all pixels and Sintel sequences in Experiment 1b. For each respective method, the parameter value selected by this procedure is used for the Middlebury sequences in Experiment 1a as well.

6.1 *Experiment 1—Data Generation*

To generate synthetic HDR sequences based on the Middlebury and Sintel datasets, we first generate downsampled and grayscale-converted images $I_f(\mathbf{x}) \in [0,1]$ that we consider to be our original images. To simulate a HDR scenario, the pixel values of I_1, I_3 are thresholded such that all pixels where $I_1(\mathbf{x}) > 0.6$, $I_3(\mathbf{x}) > 0.6$ are set to 0.6 (exposure setting I). Similarly, I_2, I_4 are thresholded by setting the pixels where $I_2(\mathbf{x}) < 0.3$, $I_4(\mathbf{x}) < 0.3$ to 0.3 (exposure setting II). Thus, the thresholded I_1, I_3 are saturated in high-intensity image regions, similarly to images taken with a long exposure setting, and are thus thought of as $I_1^{\delta t_{\mathrm{Long}}}$, $I_3^{\delta t_{\mathrm{Long}}}$. Likewise, I_2, I_4 become $I_2^{\delta t_{\mathrm{Short}}}$, $I_4^{\delta t_{\mathrm{Short}}}$. The described thresholding scheme results in sequences that contain photometrically aligned images. It is a more direct way to achieve essentially the same outcome as that of specifying some fictive raw illuminance data and exposure durations δt_{Long}, $\delta t_{\mathrm{Short}}$ for the Middlebury and Sintel frames and photometrically aligning the exposure data. The nonsaturated domain $\Omega^{\mathrm{Exp.II}}$ is determined by the nonsaturated pixels in the reference image I_2, whereas $\Omega^{\mathrm{Exp.I}}$ is determined by the points that are nonsaturated in I_3. Since the weights $\theta_{p_n q_n}(\mathbf{x})$ are expressed in reference coordinates, for methods (F) and (G), the points in $\Omega^{\mathrm{Exp.I}}$ are warped to their position in the reference frame I_2 during the estimation process using the current flow estimates. For (E), the weight $c(\mathbf{x})$ is determined using the final flow estimate $\mathbf{w}_2^{(C)}$.

6.2 *Experiment 1a—Middlebury*

The first experiment is performed on the four sequences of Middlebury that contain more than two frames as well as ground truth flow data, named Grove2, Grove3, Urban2, and Urban3. Frames 9 to 12 are taken from each sequence as $I_1, \ldots, I_4$ here. The frames from the Grove2 sequence are shown in Fig. 2. (a) shows the original images with pixel resolution 120×160. The simulated HDR sequence, generated according to Sec. 6.1 and that contains differently exposed frames, is shown in (b) and the corresponding white masks that illustrate which pixel locations are saturated in the respective frames are shown in (c). Even though a larger range of intensities, [0.6,1], are saturated in frames I_1, I_3 than the range [0,0.3] in I_2, I_4, a significantly larger number of pixels are saturated in the latter frames, since the image sequence contains predominantly low-intensity pixel values. This property is even stronger for the two Urban sequences, as can be seen for the example of Urban2 in Fig. 3, particularly by observing the vast number of saturated pixels in I_2 and I_4. All of the data costs (E), (F), (G) proposed in Sec. 5.1 use image sequences according to the alternating exposure settings described in Sec. 6.1, as depicted in Figs. 2(b) and 3(b). The methods $(A)^{\mathrm{Exp.I}}$ and $(B)^{\mathrm{Exp.I}}$, on the contrary, use input frames that are all generated according to exposure setting I, corresponding to the long exposure duration with clipped high intensity data. Similarly, the methods $(A)^{\mathrm{Exp.II}}$ and $(B)^{\mathrm{Exp.II}}$ only use input frames generated according to exposure setting II, corresponding to a short exposure duration with clipped low intensity data.

Color coded flow estimate results for the sequences Grove2 and Urban2 are displayed in Fig. 4. The text chart shows the order in which the flow estimates for the various data costs are presented and the circular color chart shows the color encoding scheme for the flow vectors in each pixel. The flow magnitude corresponding to the radius of the circle differs for each sequence, and corresponds to 1.2 times the maximum flow magnitude of the ground truth flow of a particular sequence. The maximum flow magnitudes for the Middlebury sequences used are 1.283 (Grove2), 4.812 (Grove3), 5.567 (Urban2), and 4.398 (Urban3). The average endpoint error (AEPE) over all pixels and sequences are given in Table 2, along with the value for α_S used for each method and the average angular errors (AEE)[16], as well as the AEPE obtained by employing each method on the original data (in the sense that is described in Sec. 6.1) with all weights $\theta_{p_n q_n} = 1$, $\forall \mathbf{x}$. The AEPE for each sequence is given in Table 3. The two pixels closest to the exterior of the image domain are excluded in the calculation of the AEPE and AEE scores. The flow estimates $(A)^{\mathrm{Exp.II}}$ and $(B)^{\mathrm{Exp.II}}$ are very poor for the Urban sequences, due to only images taken with exposure setting II (such as I_2 and I_4 in Fig. 3) that contain large saturated image regions. On the contrary, because the images taken with exposure setting I in Urban 2 contain very few saturated pixels, the methods $(A)^{\mathrm{Exp.I}}$ and $(B)^{\mathrm{Exp.I}}$ perform very well. In fact, they have an advantage over the methods that use alternately exposed image sequences due to a higher frame rate at their disposal which shows in their lower AEPE for Urban2. The method (G) is limited to exploiting the higher frame rate only for mutually nonsaturated regions. A temporal aspect of the proposed methods is that (F) and (G) use data from the timespan it takes to capture four images, similar to the temporally

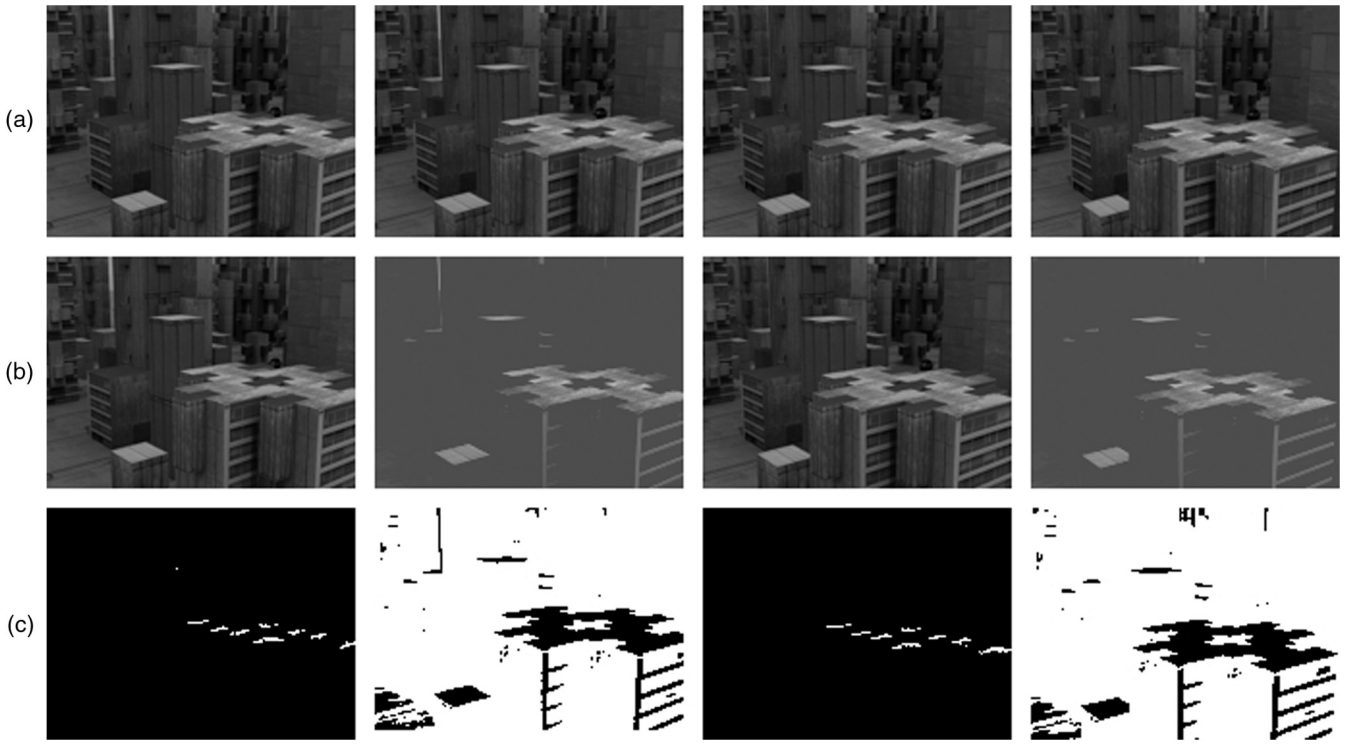

Fig. 3 The image sequence Urban2 from the Middlebury dataset: (a) original images, (b) simulated HDR sequence, used for the experimental evaluation, and (c) masks that show saturated pixels in white.

regularized baseline methods $(\mathrm{B})^{\mathrm{Exp.I}}$ and $(\mathrm{B})^{\mathrm{Exp.II}}$. Since each of the estimates used for (E) only makes use of data from the time-span that it takes to capture three images, it is conceptually closer to the methods $(\mathrm{A})^{\mathrm{Exp.I}}$ and $(\mathrm{A})^{\mathrm{Exp.II}}$ than (F) and (G) are, in the respect that less temporal regularization is included. With that reasoning, one could expect (E) to perform well and (F), (G) to perform worse, relatively speaking, for sequences where (A) performs better than (B), but such a trend is not obvious from the data in Experiment 1a. In fact, the weighted sum operation in (E) can be seen as a sort of temporal regularization itself.

6.3 *Experiment 1b—MPI Sintel*

Experiment 1b is similar to the previous Experiment 1a but is performed on the dataset of MPI Sintel, particularly on frames 11 to 14 of each sequence. An example sequence, Alley2, is shown in Fig. 5. The original (in the sense of Sec. 6.1) Alley2 frames are shown in (a). Their pixel resolution is 109×256. The differently exposed input image sequence as well as the corresponding saturation masks are shown in the Figs. 5(b) and 5(c). The estimated flow fields are given in (d) and (e). The AEPE and AEE over all pixels and all 23 sequences in the dataset are summarized in Table 4. The AEPE for each sequence are presented separately in Table 5. For the Alley2 sequence in Fig. 5, numbered here as sequence 2, the methods $(\mathrm{A})^{\mathrm{Exp.II}}$ and $(\mathrm{B})^{\mathrm{Exp.II}}$ that only use exposure setting II fail to capture the full extent of the moving person due to the saturated low-intensity data, primarily in the lower image regions. Method (E), which actually has the best AEPE for the sequence, also fails to properly estimate the motion of the moving person. In the case of method (E), the failure is due to suboptimal regularization weights for that particular sequence, unlike the case for $(\mathrm{A})^{\mathrm{Exp.II}}$ and $(\mathrm{B})^{\mathrm{Exp.II}}$. The moving person is much better captured by the estimates $(\mathrm{A})^{\mathrm{Exp.I}}$ and $(\mathrm{B})^{\mathrm{Exp.I}}$, showing that only one of the two exposure settings was sufficient for this sequence. That is, the combined dynamic range of the foreground objects (only one moving person in this sequence) did not actually demand the use of multiple exposure settings.

For some of the sequences, the flow field estimates are really poor for all methods. Due to this, the total AEPE is bloated. With that said, even the highest ranking OF methods on the MPI Sintel ranking list fails just as badly on some of the test image sequences (that are as challenging as the training images). On average, the methods $(\mathrm{A})^{\mathrm{Exp.I}}$ and $(\mathrm{B})^{\mathrm{Exp.I}}$ that only use one exposure setting perform the best in this experiment. This is due to the fact that, just as for the Urban2 sequence discussed in Experiment 1a, many of these sequences are relatively unaffected by the clipping of high-intensity data. Most often only the background regions are affected and, more so than for foreground objects, these regions are handled rather well by the regularization terms. At the bottom row of Table 5, the AEPE is presented for the case where the seven sequences whose best flow estimate has an AEPE above 1.000 are excluded. If these hard sequences are left out of the evaluation, other methods catch up in performance, indicating that the sequences where all methods more or less fail impact the AEPE of the various methods differently. To isolate some factors that influence the flow estimate results, we can first compare the results of $(\mathrm{A})^{\mathrm{Exp.I}}$ to (C) whose AEPE over all sequences is 2.241. Both methods use only two input images taken with exposure setting I. The difference is that $(\mathrm{A})^{\mathrm{Exp.I}}$ uses $I_2^{\mathrm{Exp.I}}$ and $I_3^{\mathrm{Exp.I}}$, whereas (C) uses $I_1^{\mathrm{Exp.I}}$ and $I_3^{\mathrm{Exp.I}}$. The superior

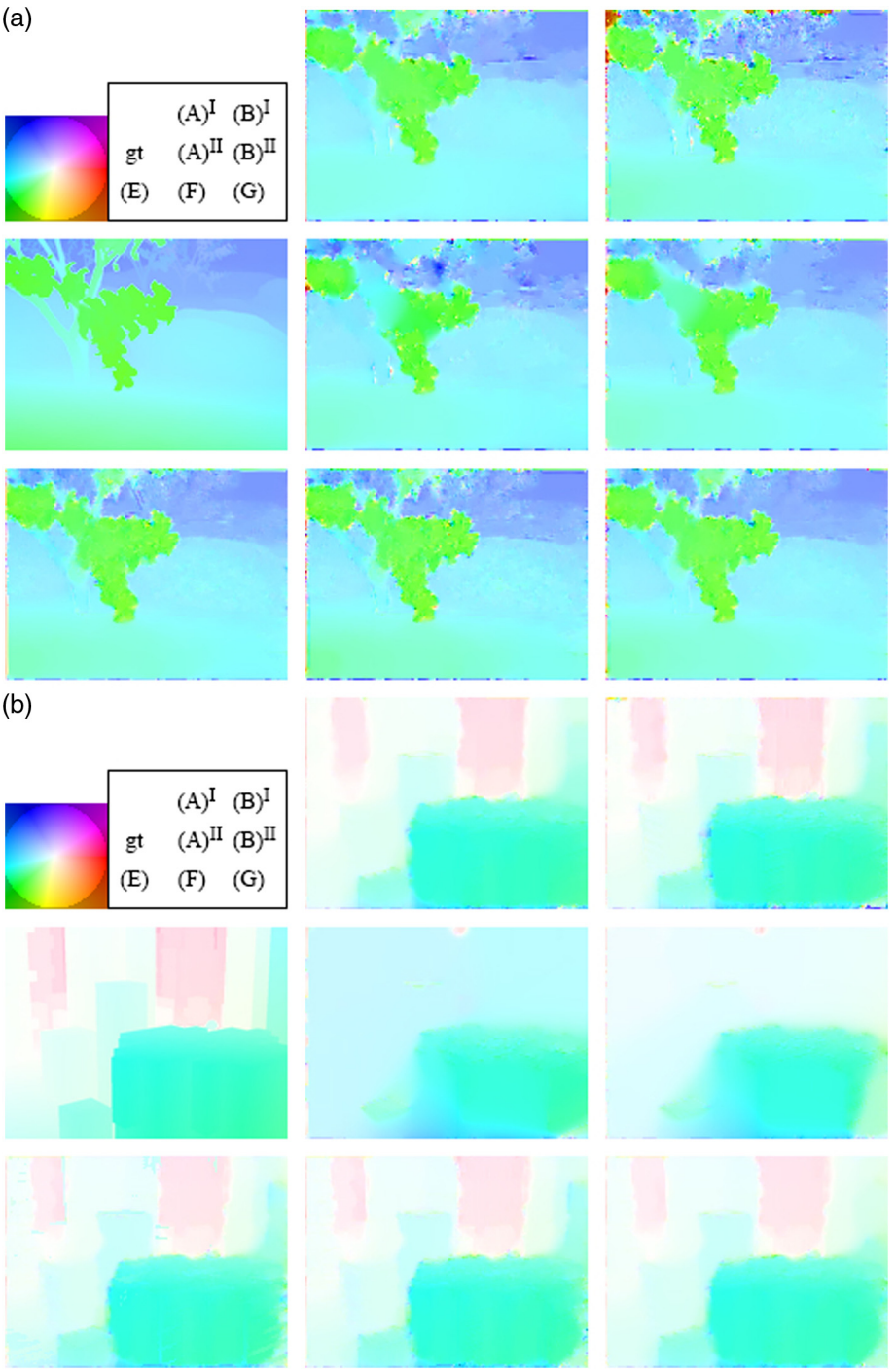

Fig. 4 Color coded flow field estimates for the various methods, encoded as shown by the circular chart. The ground truth flow is denoted "gt" in the legend. The results correspond to (a) Grove2 sequence and (b) Urban2 sequence.

AEPE of $(\mathrm{A})^{\mathrm{Exp.I}}$ is likely due to its higher frame-rate. What can be said about the proposed methods? If, e.g., (F) is compared to (C) and (D) (the latter has an AEPE of 2.696), methods that use the same frame-rate in each individual data term, (F) along with the other proposed methods show an advantageous performance. Among the set of proposed methods, there is a small benefit in performance for methods (F) and (G) that consist of minimizing a single cost functional relative to the weighted sum of (E), as well as a small benefit for (G) over (F) due to adding a data term between mutually nonsaturated regions in I_2, I_3. It is worth noting that, due to keeping α_T fixed in the parameter-selection, (B), (F), and (G) were given a slight disadvantage compared to (A) and (E), for which the value of α_T does not impact the estimation result. Furthermore, it seems that methods (B), (F), (G) more often than the other methods suffer from poor flow estimates at the edge of the image domain due to appearing or disappearing objects, which then somewhat impacts the

Table 2 Summarized results for Experiment 1a, averaged over all image sequences. For method (E), the values of α_S are. 02 for (C) and. 045 for (D).

Case	AEPE	AAE	α_S	AEPE original
$(A)^{Exp.I}$	0.231	6.86°	0.03	0.224
$(B)^{Exp.I}$	0.231	6.70°	0.025	0.225
$(A)^{Exp.II}$	0.706	19.27°	0.02	0.234
$(B)^{Exp.II}$	0.598	15.92°	0.035	0.226
(E)	0.239	6.88°	0.02/0.045	0.228
(F)	**0.230**	**6.64°**	0.03	0.217
(G)	0.233	**6.64°**	0.06	0.213

Note: Bold values indicate the best method.

Table 4 Summarized results for Experiment 1b.

Case	AEPE	AEE	α_S	AEPE original
$(A)^{Exp.I}$	**1.801**	**15.78°**	0.03	1.740
$(B)^{Exp.I}$	1.854	16.13°	0.025	1.815
$(A)^{Exp.II}$	2.331	24.07°	0.02	1.750
$(B)^{Exp.II}$	2.414	23.98°	0.035	1.817
(E)	2.185	19.03°	0.02/0.045	2.163
(F)	2.157	18.90°	0.03	2.284
(G)	2.130	17.83°	0.06	2.105

Note: bold values indicate the best method

Table 3 AEPE per image sequence for Experiment 1a. Abbreviations $(A)^{I}$, $(B)^{I}$, $(A)^{II}$, $(B)^{II}$ are used for $(A)^{Exp.I}$ and so on.

Seq.	$(A)^{I}$	$(B)^{I}$	$(A)^{II}$	$(B)^{II}$	(E)	(F)	(G)
Grove2	0.111	0.124	0.133	0.109	0.087	0.088	**0.086**
Grove3	0.264	0.287	0.282	0.268	0.251	0.262	**0.248**
Urban2	**0.198**	0.209	1.097	0.700	0.293	0.273	0.294
Urban3	0.352	0.305	1.310	1.314	0.324	**0.297**	0.302
Total	0.231	0.231	0.706	0.598	0.239	**0.230**	0.233

Note: bold values indicate the best method.

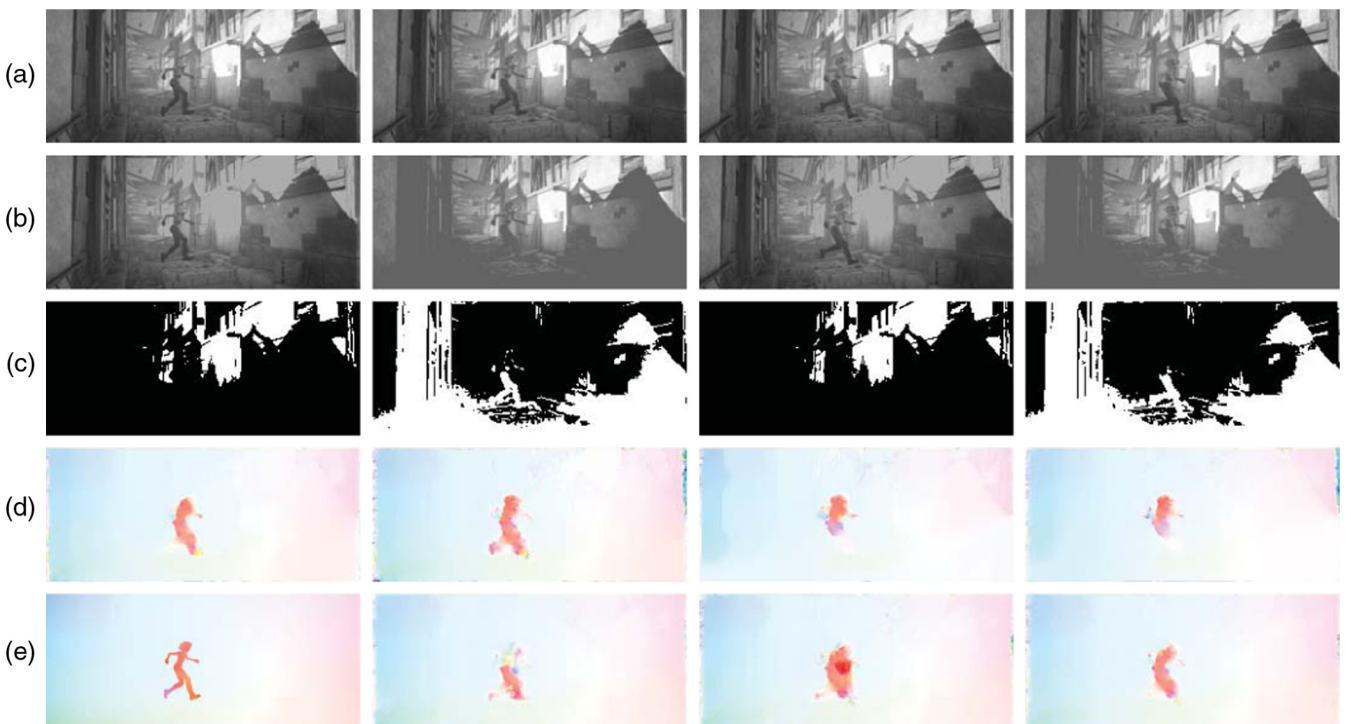

Fig. 5 (a)–(c) The image data from the MPI Sintel sequence Alley2 (Seq. # 2); (d) from left to right, flow estimates $(A)^{Exp.I}$, $(B)^{Exp.I}$, $(A)^{Exp.II}$, $(B)^{Exp.II}$; and (e) ground truth flow, flow estimates (E), (F), and (G).

Table 5 AEPE per image sequence for Experiment 1b. The total AEPE for all 23 sequences as well as for the 16 sequences with the lowest respective AEPE are given at the bottom.

Sequence no	$(A)^{I}$	$(B)^{I}$	$(A)^{II}$	$(B)^{II}$	(E)	(F)	(G)
1	0.185	0.245	0.319	0.318	0.188	0.185	**0.166**
2	0.179	0.183	0.308	0.271	**0.167**	0.199	0.170
3	9.320	**8.761**	9.921	9.444	8.956	8.923	8.949
4	6.523	6.565	**6.217**	6.253	6.976	6.975	6.781
5	1.065	1.060	1.236	1.271	0.989	**0.986**	1.114
6	5.126	6.120	**5.076**	5.450	6.808	6.381	5.896
7	**0.892**	0.905	1.677	1.647	1.118	1.167	1.039
8	0.149	0.169	0.192	0.168	**0.137**	0.154	0.138
9	0.194	0.216	0.205	0.197	0.211	0.226	**0.193**
10	**0.518**	0.521	0.639	0.607	0.546	0.550	0.531
11	**0.277**	0.300	0.406	0.411	0.317	0.318	0.301
12	**1.488**	1.705	7.008	7.019	3.960	3.925	5.241
13	1.535	**1.370**	3.135	3.183	1.972	1.939	1.934
14	0.589	0.601	0.614	**0.586**	0.690	0.744	0.701
15	**7.312**	7.916	8.685	10.753	10.239	9.929	9.436
16	**2.904**	3.059	3.867	3.892	3.775	3.937	3.444
17	0.799	0.826	**0.660**	0.666	0.783	0.795	0.728
18	**0.085**	0.087	0.299	0.304	0.114	0.114	0.114
19	0.275	**0.267**	0.570	0.563	0.288	0.285	0.274
20	0.093	0.101	0.224	0.210	0.079	0.073	**0.072**
21	0.063	0.057	0.184	0.151	**0.045**	0.048	0.047
22	0.673	**0.661**	1.154	1.113	0.873	0.733	0.703
23	1.177	**0.948**	1.029	1.049	1.019	1.031	1.013
Total	**1.801**	1.854	2.331	2.414	2.185	2.157	2.130
Best16	0.451	**0.447**	0.607	0.596	0.473	0.476	0.456

Note: bold values indicate the best method.

AEPE scores. If two additional data terms, $\theta_{12}F_{D12}$ and $\theta_{34}F_{D34}$, are added to method (G), where $\theta_{12} = \theta_{34} = \theta_{23}$ and α_S is reoptimized to 0.035, the total AEPE is decreased to 2.101. However, the improvement is due to a somewhat less bad performance on the hard sequences, as its AEPE on the best 16 sequences is 0.482, worse than all of (E), (F), and (G). If, for a change, method (F) is weighted by $\theta_{13} = \theta_{24} = 1$, $\forall \mathbf{x}$, its AEPE with re-optimized $\alpha_S = 0.025$ becomes 2.243, worse than all of the proposed methods. Overall, the results of Experiment 1b show that a method that uses alternately exposed images is not to be used as a default configuration. However, for certain imaged scenes where the dynamic range limitation is significant, a camera-mode that alternates between different exposure settings does have merit.

6.4 *Experiment 2—on Data from our Prototype Camera*

The image data in the second experimental setup comes from a camera prototype for traffic monitoring in a vehicle. Every other frame is captured with flash illumination, including light in the near-infrared spectrum, from the headlights of the car. The pixel elements of the sensor grid have two different spectral sensitivities. One quarter of the pixels are sensitive to visual light (corresponding to the typical RGB spectral bands) and the remaining pixels have a wider spectral sensitivity that includes light in the near-IR region in addition to the visual sensitivity. At each time instance, there are two image channels produced after demosaicing the sensor data. However, to conduct this experiment, a

single image channel is retained per time instance to stick with scalar valued frames in this work. Thus, data from the wideband pixels is selected. The original pixel resolution of the sensor is 720×1280. As data input $I_1, \ldots, I_4$ to the flow estimate methods, a region of size 150×250 pixels is cut out from the right part of the road in front of the vehicle on which the camera is mounted. The four input frames are displayed in Fig. 6(a). Flash illumination was used to capture I_1, I_3 while I_2, I_4 were captured without flash. In this experiment, the sampling intervals of the frames are nonuniform. For the dataset presently discussed, $t_3 - t_2 = 3(t_2 - t_1) = 3(t_4 - t_3)$. To cope with this, a minor adjustment to the regularization terms Eqs. (14) and (15) is necessary. Each flow component u_f, v_f needs to be multiplied with a constant τ_f which is inversely proportional to the time to the next sampling instance, here taken as $\tau_1 = 1$, $\tau_2 = 1/3$, $\tau_3 = 1$. This is left out of the cost functional expressions as it is a straightforward extension.

Only the methods (C), (D), (E), and (F) are considered for this image sequence. The methods that use a single exposure setting are not applicable to the available data sequences in this experiment, and the method (G) requires photometric alignment of nonsaturated regions of the differently exposed frames I_2, I_3 for which there exists no generative data model. To achieve the best possible estimates (according to manual inspection as no ground truth data is available) for the respective methods required some adjustment of the parameters. For cases (C) and (F), $\alpha_S = 0.5$ was used, whereas for method (D), $\alpha_S = 0.1$. For all cases, $\alpha_T = 0.01$. Bear in mind that the input data follow a different generative process than that of Experiment 1, discussed in Sec. 6.1. Here, linear intensity (i.e., not gamma encoded) raw sensor data with a bit depth of 12 is used. The weight functions for (E) and (F), respectively, are simply set to $c = d = 0.5$ and $\theta_{13} = \theta_{24} = 1$, $\forall \mathbf{x}$, in this experiment. Flow field estimates for Experiment 2 are shown Fig. 6(b). The dominant flow is caused by the motion of the ego vehicle in which the camera is mounted as it moves forward along the road. The

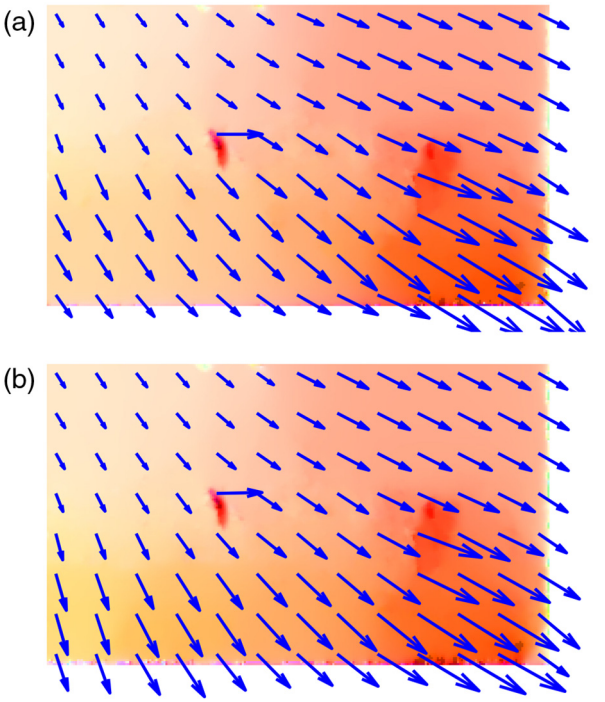

Fig. 7 (a) and (b) Quiver plots for methods (C) and (F). The magnitude of the flow vectors are scaled by a factor of 10.

flow at pixels in the right part of the image domain have a direction downward and to the right. The further left in the image domain, the more the flow direction points straight downward (while interpreting these results, remember that the used data is cropped from an originally larger image domain). A lone leg movement by one of the deer depicted in the image sequence is clearly captured by the estimated flow except for method (D). Some movement of the head of another deer stands out as well. As an additional illustration, quiver plots of the flow field estimates from cases (C)

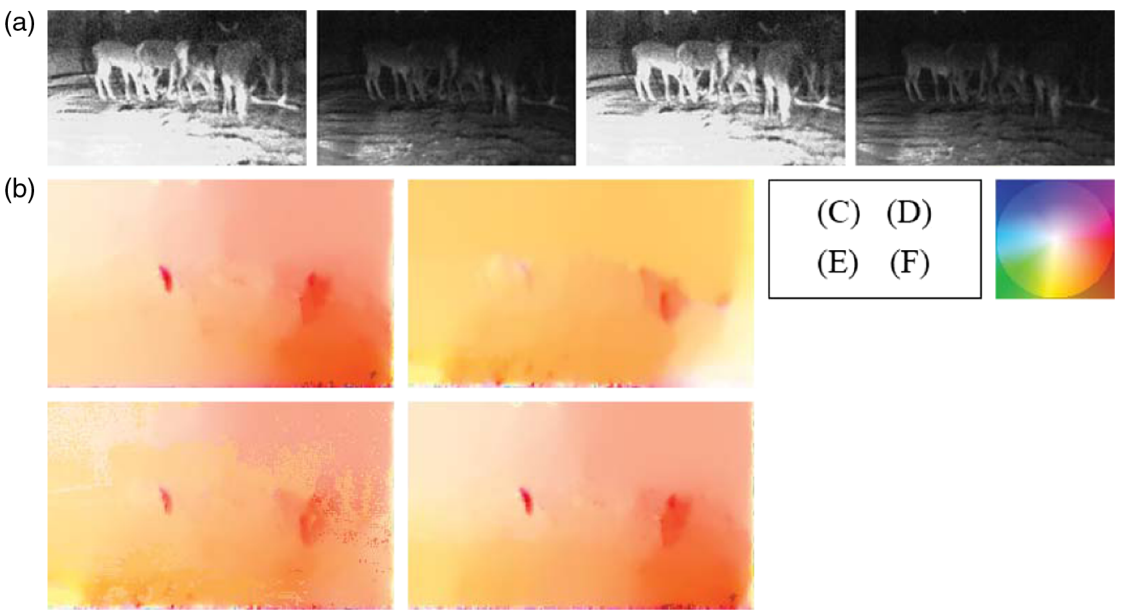

Fig. 6 Image data sequence from our prototype camera system and the flow estimates for a set of data terms.

and (F) are shown in Fig. 7. Estimated flow vectors are shown for every 20th pixel, with the flow magnitude scaled by a factor of 10 for clearer visualization. As can be seen, the direction and magnitude of the flow differ significantly between the two cases, particularly in the lower left corner, in which the road surface is mostly saturated in I_1, I_3. The improvement gained by including nonflash data I_2, I_4 in (F) compared to (C) is clear, as determined by manually inspecting the movement of the road texture between the reference frame I_2 and I_4. To round off the section, we remark that an issue with spatial regularization in OF in general is that it only measures spatial proximity in the 2-D projected image domain. In reality, objects that are closer to the camera have a larger projected motion compared to objects further away. This effect could be taken into account if depth information is available from the sensor system.

7 Conclusions and Future Work

In this work, we have considered OF estimation on image sequences with differently exposed frames. A suitable application is the case of HDR scenes where, due to the dynamic range limitation of camera sensors, each respective image inherently contains saturated regions. To address this issue, a set of data terms have been designed. These proposed methods are evaluated on synthetic datasets as well as on preliminary data from our own camera prototype. Not surprisingly, the quality of flow estimates depends directly on the quality of the input data. For example, in Experiments 1a and 1b, the methods that only use images from exposure setting II, $(\mathrm{A})^{\mathrm{Exp.II}}$ and $(\mathrm{B})^{\mathrm{Exp.II}}$, where low-intensity data are clipped generally suffer performance-wise. The proposed data terms are not designed for general image sequences, but specifically for scenarios with significant saturation in the image data for both exposure settings. As shown by the strong overall performance of $(\mathrm{A})^{\mathrm{Exp.I}}$ and $(\mathrm{B})^{\mathrm{Exp.I}}$, exposure setting I was relatively unaffected by the clipping of high-intensity data. Thus, it remains to further test our approach on real-world scenarios. No conclusive answer has been given as to which proposed method, (E), (F) or (G), is best in general. Similarly to the choice between the methods (A) and (B), which is about whether or not to use temporal regularization in the conventional OF case, the choice among the proposed methods is, to some degree, data-dependent. However, for cases where (G) is applicable, it is favorable over (F) due to its inclusion of data terms between mutually nonsaturated regions of pairs of consecutive images. Also, the presented experiments support the claim that, on average, it is favorable to use the respective data constraints in a coupled estimation approach, which is the case for (F) and (G) but not for (E). As future work, a database of HDR image sequences from, e.g., scenes that contain moving objects in both indoor and outdoor environments simultaneously is highly desired. For such a case, each respective exposure setting typically leads to significant saturation in the image regions for which it was not tuned. Additional future work includes using spatially varying weights $\theta_{p_n q_n}$ based on a specified generative data model, as well as the inclusion of feature matches and modeling of illumination-variations in the context of differently exposed image sequences.

Appendix A: Cost Functional, Corresponding Euler–Lagrange Equations and Implementation Details

In order to estimate a given flow field, the objective is to find the horizontal and vertical flow functions of $\mathbf{w}_f = (u_f, v_f)$, $\forall f \in \{1, \ldots, \mathcal{F}-1\}$ that minimize the cost functional E in Eq. (5). The flow field of interest is $\mathbf{w}_{f_{\mathrm{ref}}}$, where $f_{\mathrm{ref}} = 2$ throughout the paper. A necessary condition for a minimizer is that the first variation of E with respect to each of its arguments is equal to zero.[39] An equivalent condition is given by

$$\frac{\delta E}{\delta u_{f_0}} = 0 \Leftrightarrow \frac{\partial F}{\partial u_{f_0}} - \frac{\partial}{\partial x}\frac{\partial F}{\partial u_{f_0 x}} - \frac{\partial}{\partial y}\frac{\partial F}{\partial u_{f_0 y}} = 0, \quad \forall\, f_0, \tag{16}$$

$$\frac{\delta E}{\delta v_{f_0}} = 0 \Leftrightarrow \frac{\partial F}{\partial v_{f_0}} - \frac{\partial}{\partial x}\frac{\partial F}{\partial v_{f_0 x}} - \frac{\partial}{\partial y}\frac{\partial F}{\partial v_{f_0 y}} = 0, \quad \forall\, f_0, \tag{17}$$

where the left hand sides are the first variations of the functional E with respect to u_{f_0} and v_{f_0}, $f_0 \in \{1, \ldots, \mathcal{F}-1\}$, and the right hand sides are the (strong form) E–L partial differential equations, for which each of the terms u_{f_0}, $u_{f_0 x}$, $u_{f_0 y}$, v_{f_0}, $v_{f_0 x}$, $v_{f_0 y}$ are treated as independent variables. Due to the equivalences in Eq. (16), the flow that minimizes the total cost functional is obtained by the solution to the E–L equations. With $F = F_D + \alpha_S F_S + \alpha_T F_T$, we make use of the linearity of the derivative operator to evaluate the respective terms

$$F_D = \sum_{n=1}^{N} \theta_{p_n q_n} F_{D p_n q_n},$$
$$F_{D p_n q_n} = \Psi\{[I_{q_n}(\mathbf{x} + \mathbf{W}_{q_n}) - I_{p_n}(\mathbf{x} + \mathbf{W}_{p_n})]^2\}, \tag{18}$$

$$F_S = \Psi\left[\sum_{f=1}^{\mathcal{F}-1}(u_{fx}^2 + u_{fy}^2 + v_{fx}^2 + v_{fy}^2)\right], \tag{19}$$

$$F_T = \sum_{f=1}^{\mathcal{F}-2} \Psi[(u_{f+1} - u_f)^2 + (v_{f+1} - v_f)^2], \tag{20}$$

part by part for F in Eq. (16) in the following sections of the appendix. The contributions are added together at the end. Note that F_S only depends on the derivatives of u_{f0}, v_{f0} and that F_D, F_T only depend on u_{f0}, v_{f0} themselves. Thus, for each of F_D, F_S, F_T, some of the terms in Eq. (16) vanish directly, simplifying the respective evaluations. All the derivations are given for Eq. (16), with respect to u_{f_0} (for a specific f_0). The derivation of the E–L equations in (17) with respect to v_{f_0} is analogous. Due to the general choices allowed for p_n, q_n, and f_{ref}, and the fact that the expression is nonlinear in the unknown flow, the evaluation of the F_D part is somewhat cumbersome. The regularization terms are relatively straightforward to treat, and are, therefore, given first. If the interval between the sampling instances is not 1, or particularly if the sampling intervals are nonuniform which is the case for the camera prototype setup in Experiment 2, a scale factor is necessary for each flow

increment. While it is included in our implementation, it is left out of the derivation as it is a straightforward extension.

A.1 *Spatial Regularization Term*

The contribution from the spatial term Eq. (19) to the E–L equations of (16) is given by

$$-\frac{\partial}{\partial x}\frac{\partial F_S}{\partial u_{f_0x}}-\frac{\partial}{\partial y}\frac{\partial F_S}{\partial u_{f_0y}}=-2\,\mathrm{div}(\Psi'_S\nabla u_{f_0}),$$
$$\Psi'_S\triangleq\Psi'\left[\sum_{f=1}^{\mathcal{F}-1}(u_{fx}^2+u_{fy}^2+v_{fx}^2+v_{fy}^2)\right], \tag{21}$$

where $\Psi'(z^2)=(1/2)(z^2+\epsilon^2)^{-1/2}$, due to

$$\frac{\partial F_S}{\partial u_{f_0}}=0,\quad \frac{\partial}{\partial x}\frac{\partial F_S}{\partial u_{f_0x}}=\frac{\partial}{\partial x}(\Psi'_S 2u_{f_0x}),$$
$$\frac{\partial}{\partial y}\frac{\partial F_S}{\partial u_{f_0y}}=\frac{\partial}{\partial y}(\Psi'_S 2u_{f_0x}). \tag{22}$$

Although short notation is used, the reader is reminded that u_{f_0x} and u_{f_0y} are the functions of (x,y). An interesting observation is that the contribution Eq. (21) of the spatial term to the E–L equations has the form of a nonlinear (due to the dependence of Ψ'_S on the unknown parameters) diffusion, commonly used for edge-preserving image denoising.[44,45] Here, however, this term is balanced against the other included terms.

A.2 *Temporal Regularization Term*

Because the temporal regularization term Eq. (20) does not contain any partial derivatives of the flow functions in its expression, the contribution to Eq. (16) is given directly as

$$\frac{\partial F_T}{\partial u_{f_0}}=\begin{cases}\Psi'_{TI}\cdot 2(u_2-u_1)\cdot(-1), & f_0=1,\\ \Psi'_{TII}\cdot 2(u_{f_0}-u_{f_0-1})\cdot(+1)+ & \\ +\Psi'_{TI}\cdot 2(u_{f_0+1}-u_{f_0})\cdot(-1), & 1<f_0<\mathcal{F}-1,\\ \Psi'_{TII}\cdot 2(u_{\mathcal{F}-1}-u_{\mathcal{F}-2})\cdot(+1), & f_0=\mathcal{F}-1,\end{cases}$$
$$\Psi'_{TI}\triangleq\Psi'[(u_{f_0+1}-u_{f_0})^2+(v_{f_0+1}-v_{f_0})^2],$$
$$\Psi'_{TII}\triangleq\Psi'[(u_{f_0}-u_{f_0-1})^2+(v_{f_0}-v_{f_0-1})^2]. \tag{23}$$

A.3 *Data Term*

Similarly to the temporal regularization term, the data term Eq. (18) does not contain any partial derivatives of the flow in its expression, thus its contribution to Eq. (16) is

$$\frac{\partial F_D}{\partial u_{f_0}}=\sum_{n=1}^{N}\theta_{p_nq_n}\frac{\partial F_{Dp_nq_n}}{\partial u_{f_0}}, \tag{24}$$

where for each specific n,

$$\frac{\partial F_{Dp_nq_n}}{\partial u_{f_0}}=\Psi'[(I_{p_nq_n})^2]\cdot 2I_{p_nq_n}\cdot\frac{\partial I_{p_nq_n}}{\partial u_{f_0}} \tag{25}$$

and where $I_{p_nq_n}\triangleq I_{q_n}(\mathbf{x}+\mathbf{W}_{q_n})-I_{p_n}(\mathbf{x}+\mathbf{W}_{p_n})$. To further evaluate Eq. (25), which is nonlinear in the flow functions u_{f_0} and v_{f_0} contained in $\mathbf{W}_{q_n}$ and $\mathbf{W}_{p_n}$, successive linearizations about the current flow estimates are employed in an iterative scheme. This is what is referred to as a warping scheme in the papers cited in the introduction of the original paper.[3] The warping scheme typically relies on a coarse-to-fine multiresolution strategy to avoid local minima, with implications discussed in the introduction. The flow functions are separated into the current estimate at iteration (k) and a flow update term, according to

$$\mathbf{w}_f\to\mathbf{w}_f^{(k+1)}=\mathbf{w}_f^{(k)}+\mathbf{dw}_f^{(k)}, \tag{26}$$

such that

$$I_{q_n}(\mathbf{x}+\mathbf{W}_{q_n})\to I_{q_n}(\mathbf{x}+\mathbf{W}_{q_n}^{(k+1)})\approx$$
$$\approx I_{q_n}(\mathbf{x}+\mathbf{W}_{q_n}^{(k)})+I_{q_nx}^{(k)}dU_{q_n}^{(k)}+I_{q_ny}^{(k)}dV_{q_n}^{(k)},$$
$$I_{p_n}(\mathbf{x}+\mathbf{W}_{p_n})\to I_{p_n}(\mathbf{x}+\mathbf{W}_{p_n}^{(k+1)})\approx$$
$$\approx I_{p_n}(\mathbf{x}+\mathbf{W}_{p_n}^{(k)})+I_{p_nx}^{(k)}dU_{p_n}^{(k)}+I_{p_ny}^{(k)}dV_{p_n}^{(k)}, \tag{27}$$

where

$$I_{q_nx}^{(k)}=\tfrac{\partial}{\partial x}[I_{q_n}(\mathbf{x}+\mathbf{W}_{q_n}^{(k)})],\quad I_{q_ny}^{(k)}=\tfrac{\partial}{\partial y}[I_{q_n}(\mathbf{x}+\mathbf{W}_{q_n}^{(k)})], \tag{28}$$

and similarly for $I_{p_nx}^{(k)}$, $I_{p_ny}^{(k)}$. Expressions of the type $I_{q_n}(\mathbf{x}+\mathbf{W}_{q_n}^{(k)})$ are computed by means of bicubic interpolation. Using the substitution in Eq. (27), we get for Eq. (25) that

$$\frac{\partial I_{p_nq_n}^{(k)}}{\partial u_{f_0}}=\begin{cases}I_{q_nx}^{(k)}-I_{p_nx}^{(k)}, & f_0\geq r,\ q_n>f_0,\ p_n>f_0,\\ I_{q_nx}^{(k)}, & f_0\geq r,\ q_n>f_0,\ p_n\leq f_0,\\ I_{p_nx}^{(k)}-I_{q_nx}^{(k)}, & f_0<r,\ p_n\leq f_0,\ q_n\leq f_0,\\ I_{p_nx}^{(k)}, & f_0<r,\ p_n\leq f_0,\ q_n>f_0,\\ 0, & \text{else},\end{cases} \tag{29}$$

where $q_n>p_n$, $\forall n$, by construction. For $\partial I_{p_nq_n}^{(k)}/\partial v_{f_0}$ in the E–L Eq. (17) with respect to. v_{f_0}, all partial derivatives of Eq. (29) are with respect to y instead of x.

A.4 *Pseudoalgorithm for the Minimization Procedure*

The full E–L equations stated in Eq. (16) are given by summing the contributions Eqs. (21), (23), and (25) together (using the weights α_S, α_T). The flow terms in all expressions are replaced, for the sake of the iterative solution scheme, by a current estimate and an update term according to Eq. (26), as shown for the data term in the previous section. However, there still remains a nonlinearity in the E–L equations, due to the expression of Ψ', that should be dealt with. In order to obtain a linear expression of the E–L equations, an inner iteration loop over iteration index l is added in which the Ψ'-terms in Eqs. (21), (23), and (25) lag behind the flow update terms $\mathbf{dw}_f^{(k,l+1)}=(du_f^{(k,l+1)},dv_f^{(k,l+1)})$. The notation $\Psi'\{\cdot\}^{(k,l)}$ is thus introduced to refer to any of the Ψ'-terms with the flow update terms in its argument taken from the previous iteration (l), i.e., $[du_f^{(k,l)},dv_f^{(k,l)}]$. Several papers take a similar iterative approach, with an outer and an inner loop to successively linearize the problem. We refrain from typing out the full linearized E–L equations, and

suggest a study of other publications where the final expressions are not as involved, e.g., by Brox et al.[3] At a given iteration, (k,l), a set of linear equations that interconnects all E–L equations is formed and solved numerically on the pixel grid of the reference image to yield discrete approximations of $[du_f^{(k,l+1)}, dv_f^{(k,l+1)}]$, $\forall f$. The pseudoalgorithm for the full iterative scheme is given in Table 6.

A coarse-to-fine estimation strategy is used in the experiments, with $S = 10$ scale levels and a re-sampling factor of 0.85. The first scale $s = 1$ is resampled by a factor $0.85^{(10-1)} \approx 0.23$, which represents the coarsest resolution. For each scale, the algorithm in Table 6 is run, although for $s > 1$, $\mathbf{u}^{(0)}$, $\mathbf{v}^{(0)}$ are assigned values corresponding to the re-scaled solution from the previous scale. The number of outer and inner iterations used are $K = 5$ and $L = 5$, respectively, verified to be sufficient for convergence. The flow updates at each step (k,l) are computed in closed-form in the experiments, but would also demand an iterative approach (e.g., Gauss–Seidel type methods are used by other authors) for larger image resolutions. The boldface vectors $\mathbf{u}^{(k)} = [(\mathbf{u}_1^{(k)})^T, \ldots, (\mathbf{u}_{\mathcal{F}-1}^{(k)})]^T$ and $\mathbf{v}^{(k)} = [(\mathbf{v}_1^{(k)})^T, \ldots, (\mathbf{v}_{\mathcal{F}-1}^{(k)})]^T$ in Table 6 contain all flow increments $\mathbf{u}_f$ and $\mathbf{v}_f$, each of size $M \times 1$, that in turn contain flow data from the discretized image domain in vectorized form. Thus, the dimension of the equation system that results from numerically implementing the E–L equations on the pixel grid of the reference frame is $2(\mathcal{F}-1)\cdot M$, where M is the number of pixels per frame. The update terms $\mathbf{du}^{(k,l+1)}$, $\mathbf{dv}^{(k,l+1)}$ are formed similarly. All the first order derivatives are implemented with the discrete convolution kernels $[0.5, 0, -0.5]$ and $[0.5, 0, -0.5]^T$ for the horizontal and vertical cases, respectively. No prior low-pass filtering of the image at the given resolution scale is performed for the discrete derivative approximations. A reservation is made for the implementation of the divergence of the scaled gradient in Eq. (21), which has the form

Table 6 Optical flow estimation by finding the (discretized) minimizer of a variational cost functional as the solution to the corresponding Euler–Lagrange equations.

Pseudoalgorithm

initialization: $\mathbf{u}^{(0)} = 0$, $\mathbf{v}^{(0)} = 0$

for $k = 0, \ldots, K-1$

compute $I_{q_n}^{(k)}$, $I_{q_nx}^{(k)}$, $I_{q_ny}^{(k)}$, $I_{p_n}^{(k)}$, $I_{p_nx}^{(k)}$, $I_{p_ny}^{(k)}$, $\forall n$

$\mathbf{du}^{(k,0)} = 0$, $\mathbf{dv}^{(k,0)} = 0$

for $l = 0, \ldots, L-1$

compute the $\Psi'\{\cdot\}^{(k,l)}$-terms in Eqs. (18), (20), (22), $\forall n$, f_0

construct $\mathbf{A}^{(k,l)}$, $\mathbf{b}^{(k,l)}$

solve $\mathbf{A}^{(k,l)}[(\mathbf{du}^{(k,l+1)})^T, (\mathbf{dv}^{(k,l+1)})^T]^T = \mathbf{b}^{(k,l)}$

end

$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \mathbf{du}^{(k,L)}$, $\mathbf{v}^{(k+1)} = \mathbf{v}^{(k)} + \mathbf{dv}^{(k,L)}$

end

$$\operatorname{div}(\Psi_S' \nabla u_{f_0}) = \frac{\partial}{\partial x}\left(\Psi_S' \frac{\partial u_{f_0}}{\partial x}\right) + \frac{\partial}{\partial y}\left(\Psi_S' \frac{\partial u_{f_0}}{\partial y}\right) \tag{30}$$

and is implemented with convolution kernels $[1, -1]$ and $[1, -1]^T$ for the respective partial derivatives. This choice coincides with a previously proposed implementation from a similar OF method with a nonvariational formulation, where the term corresponding to the expression in Eq. (30) is referred to as a generalized Laplacian.[46] The discretization leads to the following approximation

$$\left.\frac{\partial}{\partial x}\left(g\frac{\partial u}{\partial x}\right)\right|_{i,j} \approx g_{i,j}(u_{i,j+1} - u_{i,j}) - g_{i,j-1}(u_{i,j} - u_{i,j-1}), \tag{31}$$

about a pixel (i,j) of a flow vector $\mathbf{u}$, where g is seen to be evaluated asymmetrically, yet gives convincing results in empirical tests against other discrete operators for the given experiments. The formulation in Eq. (31) allows for comparison with the rich research results on numerics of nonlinear diffusion, to which the interested reader is referred for a more thorough study.[47]

Acknowledgments

This work is funded by the Swedish research agency VINNOVA under project 2013-04702, and by Volvo Cars.

References

1. B. Horn and B. Schunck, "Determining optical flow," *Artif. Intell.* **17**(1), 185–203 (1981).
2. J. Barron, D. Fleet, and S. Beauchemin, "Performance of optical flow techniques," *Int. J. Comput. Vision* **12**(1), 43–77 (1994).
3. T. Brox et al., "High accuracy optical flow estimation based on a theory for warping," in *European Conf. on Comp. Vision (ECCV)*, pp. 25–36, Springer, Berlin Heidelberg (2004).
4. A. Wedel and D. Cremers, *Stereo Scene Flow for 3D Motion Analysis*, Springer, London (2011).
5. D. Cremers and S. Soatto, "Motion competition: a variational approach to piecewise parametric motion segmentation," *Int. J. Comput. Vision* **62**(3), 249–265 (2005).
6. W. Crum, T. Hartkens, and D. Hill, "Non-rigid image registration: theory and practice," *BJR* **77**, S140–S153 (2004).
7. D. Geronimo et al., "Survey of pedestrian detection for advanced driver assistance systems," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(7), 1239–1258 (2010).
8. E. Reinhard et al., *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*, Morgan Kaufmann, San Francisco (2010).
9. P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. 24th Annual Conf. on Computer Graphics and Interactive Techniques, SIGGRAPH '97*, pp. 369–378, ACM Press/Addison-Wesley Publishing Co., New York (1997).
10. D. Sun, S. Roth, and M. Black, "Secrets of optical flow estimation and their principles," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 2432–2439 (2010).
11. C. Vogel, S. Roth, and K. Schindler, "An evaluation of data costs for optical flow," in *German Conf. on Pattern Recognition (GPCR)*, pp. 343–353, Springer (2013).
12. O. Demetz et al., "Learning brightness transfer functions for the joint recovery of illumination changes and optical flow," in *Computer Vision-ECCV 2014*, pp. 455–471, Springer (2014).
13. B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. on Artificial Intelligence (IJCAI'81)*, pp. 674–679 (1981).
14. A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/Kanade meets Horn/Schunck: combining local and global optic flow methods," *Int. J. Comput. Vision* **61**(3), 211–231 (2005).
15. J. Braux-Zin, R. Dupont, and A. Bartoli, "A general dense image matching framework combining direct and feature-based costs," in *IEEE Int. Conf. on Computer Vision (ICCV)*, pp. 185–192 (2013).
16. S. Baker et al., "A database and evaluation methodology for optical flow," *Int. J. Comput. Vision* **92**(1), 1–31 (2011).

17. D. Butler et al., "A naturalistic open source movie for optical flow evaluation," in *European Conf. on Computer Vision (ECCV)*, pp. 611–625, Springer (2012).
18. S. Baker et al., "The Middlebury computer vision pages, optical flow page," http://vision.middlebury.edu/flow/eval/, (18 August 2015).
19. Butler D. et al., "MPI sintel flow dataset," http://sintel.is.tue.mpg.de/, (18 August 2015).
20. E. Memin and P. Perez, "A multigrid approach for hierarchical motion estimation," in *IEEE Int. Conf. on Computer vision (ICCV)*, pp. 933–938 (1998).
21. A. Bruhn et al., "A multigrid platform for real-time motion computation with discontinuity-preserving variational methods," *Int. J. Comput. Vision* **70**(3), 257–277 (2006).
22. F. Steinbrucker, T. Pock, and D. Cremers, "Large displacement optical flow computation without warping," in *IEEE Int. Conf. on Computer vision (ICCV)*, pp. 1609–1614 (2009).
23. T. Brox and J. Malik, "Large displacement optical flow: descriptor matching in variational motion estimation," *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(3), 500–513 (2011).
24. L. Xu, J. Jia, and Y. Matsushita, "Motion detail preserving optical flow estimation," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(9), 1744–1757 (2012).
25. P. Weinzaepfel et al., "Deepflow: large displacement optical flow with deep matching," in *IEEE Int. Conf. on Computer Vision (ICCV)* (2013).
26. M. Stoll, S. Volz, and A. Bruhn, "Adaptive integration of feature matches into variational optical flow methods," in *Asian Conf. on Computer Vision (ACCV)*, Vol. **7726**, pp. 1–14, Springer, Berlin Heidelberg (2013).
27. M. Leordeanu, A. Zanfir, and C. Sminchisescu, "Locally affine sparse-to-dense matching for motion and occlusion estimation," in *IEEE Int. Conf. on Computer Vision (ICCV)* (2013).
28. J. Revaud et al., "Epicflow: edge-preserving interpolation of correspondences for optical flow," arXiv preprint arXiv:1501.02565 (2015).
29. J. Weickert and C. Schnörr, "Variational optic flow computation with a spatio-temporal smoothness constraint," *J. Math. Imaging Vision* **14**(3), 245–255 (2001).
30. A. Salgado and J. Sánchez, "Temporal constraints in large optical flow estimation," in *Computer Aided Systems Theory (EUROCAST)*, pp. 709–716, Springer, Berlin Heidelberg (2007).
31. S. Volz et al., "Modeling temporal coherence for optical flow," in *IEEE Int. Conf. on Computer Vision (ICCV)*, pp. 1116–1123 (2011).
32. A. Wedel et al., "An improved algorithm for TV-L1 optical flow," in *Statistical and Geometrical Approaches to Visual Motion Analysis*, pp. 23–45, Springer, Berlin Heidelberg (2009).
33. T. Müller et al., "Illumination-robust dense optical flow using census signatures," in *Pattern Recognition*, pp. 236–245, Springer, Berlin Heidelberg (2011).
34. A. Sellent et al., "Motion field estimation from alternate exposure images," *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(8), 1577–1589 (2011).
35. D. Hafner, O. Demetz, and J. Weickert, "Simultaneous HDR and optic flow computation," in *Int. Conf. on Pattern Recognition (ICPR)*, pp. 2065–2070, IEEE (2014).
36. E. Allen and S. Triantaphillidou, *The Manual of Photography*, Focal Press (2011).
37. M. Anderson et al., "Proposal for a standard default color space for the internet–sRGB," in *Color and Imaging Conf.*, pp. 238–245, Society for Imaging Science and Technology (1996).
38. International Color Consortium, "Specification of sRGB," 2015, Reston, VA, http://www.color.org/sRGB.pdf, (18 August 2015).
39. G. Strang, *Computational Science and Engineering*, Wellesley-Cambridge Press (2007).
40. A. Chambolle and P. Lions, "Image recovery via total variation minimization and related problems," *Numerische Mathematik* **76**(2), 167–188 (1997).
41. M. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Comput. Vision Image Understanding* **63**(1), 75–104 (1996).
42. P. Blomgren and T. Chan, "Color TV: total variation methods for restoration of vector-valued images," *IEEE Trans. Image Process.* **7**(3), 304–309 (1998).
43. H. Zimmer et al., "Complementary optic flow," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp. 207–220, Springer (2009).
44. J. Weickert and T. Brox, "Diffusion and regularization of vector-and matrix-valued images," *Contemp. Math.* **313**, 251–268 (2002).
45. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(7), 629–639 (1990).
46. C. Liu, "Beyond pixels: exploring new representations and applications for motion analysis," PhD Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (2009).
47. T. Brox, "From pixels to regions: partial differential equations in image analysis," PhD Thesis, Department of Mathematics and Computer Science, Saarland University (2005).

Tomas Bengtsson is a PhD student at Chalmers University of Technology, where he received his BS degree in electrical engineering and his MS degree in communication engineering in 2007 and 2009, respectively. His current research focuses on motion estimation from camera data. He has formerly published research in SPIE on super-resolution and high dynamic range image reconstruction.

Tomas McKelvey received his PhD degree in automatic control at Linköping University in 1995. Between 1995 and 1999 he held research and teaching positions at Linköping University. Since 2000 he has been with Chalmers University of Technology and since 2006, he has held a full professor position in signal processing. His research interests are model-based signal processing, system identification, and automatic control with applications to biomedical engineering combustion engines, and automotive active safety.

Konstantin Lindström received his BSc in physics in 1996 and his MSc in mathematics in 1997 from Gothenburg University in Sweden. He has since then been conducting research and consulting for the defense and automotive industries. He is currently employed at Volvo Car Corporation as a strategic data analyst.