

Learning Brightness Transfer Functions for the Joint Recovery of Illumination Changes and Optical Flow

Oliver Demetz^{1,*}, Michael Stoll^{2,*}, Sebastian Volz²,
Joachim Weickert¹, and Andrés Bruhn²

¹ Mathematical Analysis Group, Saarland University, Saarbrücken, Germany
{demetz, weickert}@mia.uni-saarland.de

² Institute for Visualization and Interactive Systems, University of Stuttgart, Germany
{stoll, volz, bruhn}@vis.uni-stuttgart.de

Abstract. The increasing importance of outdoor applications such as driver assistance systems or video surveillance tasks has recently triggered the development of optical flow methods that aim at performing robustly under uncontrolled illumination. Most of these methods are based on patch-based features such as the normalized cross correlation, the census transform or the rank transform. They achieve their robustness by locally discarding both absolute brightness and contrast. In this paper, we follow an alternative strategy: Instead of discarding potentially important image information, we propose a novel variational model that jointly estimates *both* illumination changes *and* optical flow. The key idea is to parametrize the illumination changes in terms of basis functions that are learned from training data. While such basis functions allow for a meaningful representation of illumination effects, they also help to distinguish real illumination changes from motion-induced brightness variations if supplemented by additional smoothness constraints. Experiments on the KITTI benchmark show the clear benefits of our approach. They do not only demonstrate that it is possible to obtain meaningful basis functions, they also show state-of-the-art results for robust optical flow estimation.

1 Introduction

Three decades after the seminal work of Horn and Schunck [22], dense variational optical flow methods have found their way into numerous real-world applications such as driver assistance systems [33], markerless motion capture [13], long-term trajectory analysis [32] as well as motion-aware video editing [37]. Based on the minimization of a global energy functional that combines constancy assumptions (data term) with regularity constraints (smoothness term), variational methods allow both a transparent modeling and an accurate estimation of the results. Since many real-world applications require to process outdoor sequences, it is not surprising that the robustness of optical flow methods under uncontrolled illumination has become a major challenge. This is also reflected in the design of recent real-world benchmarks such as the KITTI Vision Benchmark Suite [14]. It provides challenging data from automotive scenarios that contains typical illumination changes due to automatic camera re-adjustments, changeable weather conditions or physical effects such as shadows and highlights.

* Authors have equally contributed to this work.

Basic Optical Flow Approaches. In order to tackle the problem of illumination changes, most approaches from the literature make use of constancy assumptions based on *illumination-invariant* image features. Such features achieve their robustness by locally discarding illumination-sensitive information such as absolute brightness or contrast [24,27,34,36,50]. In the extreme case almost all information is discarded and only a relative local ordering is stored [5,11,35]. Methods based on such illumination-invariant features use constancy assumptions on higher order derivatives such as the gradient or the Hessian [23,34,38,43,51], photometric invariants [27,45,54] as well as mutual information [21]. Moreover, recently, patch-based techniques have become very popular such as the normalized cross correlation (NCC) [50], the rank transform [11,53] and the census transform [5,35,39]. Some of the approaches also use illumination-robust descriptors from sparse feature matching such as SIFT [24,25] and HOG [8,36]. A comparison of some of these methods can be found in [40] and [46]. Similar in spirit are methods that discard illumination-relevant information via preprocessing. Typically, they employ the structure-texture-decomposition [48] or derivative-type filters [41].

What all those aforementioned methods have in common is that they discard potentially valuable image information. However, if illumination changes are only moderate or not even present, discarding brightness and contrast information may significantly deteriorate the results. Moreover, some of the transformations are highly non-linear and even lack differentiability which results in a more complex optimization. Finally, most invariants are not defined at all locations, since illumination-invariant information cannot be extracted everywhere (e.g. in homogeneous regions). Summarizing: Instead of discarding potentially valuable information that may harm the estimation, it would be desirable to keep and exploit all available information when estimating the flow.

Advanced Optical Flow Approaches. In fact, there are a few methods in the literature that follow the above mentioned idea by *jointly* estimating both illumination changes and the optical flow. On the one hand, there are approaches that seek to estimate a single *global* brightness transfer function to identify problematic image regions [10]. On the other hand, there are techniques that embed the classical brightness constancy assumption into a parametrized *local* illumination model for which the coefficients are jointly estimated. Such local models include simple additive terms [7,28], affine illumination models [15,18,30] as well as complex brightness models derived from physics [20]. Recently, also local and global ideas were combined [18]: While a local affine model allows to estimate the correspondences in a PatchMatch-like approach [1], the information is eventually condensed to a single global transfer function.

Preserving potentially valuable image information, however, is not the only advantage of approaches based on parametrized illumination models. If appropriate models are employed, imposing smoothness on the resulting parameter field allows to separate real illumination changes from motion-induced brightness variations. However, so far in the literature, the considered illumination models were either chosen ad-hoc [7,15], or specifically tailored towards a certain physical process [20]. There have been no efforts so far to determine the most suitable model for a specific type of data. Moreover, it has not yet been investigated how the smoothness term of the coefficient field should be modeled such that it allows for a good separation of motion and illumination effects. Finally, since all existing variational methods with parametrized illumination models

are based on simple concepts for data and smoothness terms, it remains unclear how a more sophisticated joint method would perform on a suitable optical flow benchmark.

Our Contribution. In this paper, we address all these questions. Firstly, we use a principal components analysis (PCA)-based approach with a clustering step to learn a suitable basis for the local brightness transfer functions from training data. Secondly, we propose a joint complementary regularizer for the basis coefficients that is based on a weighting scheme derived from the eigenvalues of the PCA. Thirdly, we embed the basis functions and the regularizer into a variational model that combines brightness and gradient constancy with a second-order smoothness term. Experiments demonstrate that our approach works very well in practice. We obtain meaningful basis functions, intuitive coefficient fields as well as state-of-the-art results on the KITTI benchmark.

Related Work on Basis Learning. Apart from the aforementioned techniques that jointly estimate illumination changes and optical flow, a few more related works are worth mentioning. On the one hand, there are methods that address the estimation of camera response or brightness transfer functions, mainly in the context of HDR imaging. Such approaches include the work of Grossberg and Nayar [16] who proposed to compute the brightness transfer function via histogram specification, as well as the papers of Debevec and Malik [9] and Grossberg and Nayar [17] on estimating the camera response function, the latter one using learned basis functions. On the other hand, there are approaches that represent appearance changes with basis functions for illumination changes. Such methods include the template tracking approach of Hager and Belhumeur [19] as well as the work on iconic changes by Black *et al.* [3]. Finally, there exist a few optical flow methods that make use of spatial or temporal basis functions to model the flow. This applies to the approaches by Nir *et al.* [31] on over-parametrized optical flow and Garg *et al.* on temporal tracking of non-rigid objects with subspace constraints [12].

Organization. In Section 2, we introduce our novel variational model based on brightness transfer basis functions and joint complementary coefficient regularization. Minimization issues are then discussed in Section 3. The estimation of the basis functions and the clustering step are explained in Section 4. Our results and a comparison to the literature are presented in Section 5. The paper ends with a summary in Section 6.

2 Variational Model

Let us consider a sequence of two images $f_i : \Omega \rightarrow \mathbb{R}$ ($i \in \{1, 2\}$) defined on a rectangular domain $\Omega \subset \mathbb{R}^2$. Furthermore, let the optical flow field be denoted by $\mathbf{w} = (u, v)^\top : \Omega \rightarrow \mathbb{R}^2$ and let the illumination changes be parametrized by a coefficient field $\mathbf{c} : \Omega \rightarrow \mathbb{R}^n$. Then, inspired by the basic approach of Cornelius and Kanade [7], we propose to jointly compute the optical flow and the illumination changes as minimizer of an energy functional with the following structure:

$$E(\mathbf{w}, \mathbf{c}) = \int_{\Omega} (D(\mathbf{w}, \mathbf{c}) + R_{\text{flow}}(\mathbf{w}) + R_{\text{illum}}(\mathbf{c})) \, d\mathbf{x}. \quad (1)$$

It consists of three terms: a data term D that relates two consecutive frames of the input image sequence via the optical flow and the parametrized illumination changes (in terms

of coefficients), a flow regularization term R_{flow} which encourages a piecewise affine flow field, and a coefficient regularization term R_{illum} that assumes the coefficient fields to be piecewise smooth. Let us now discuss these terms in detail.

2.1 Data Term

Unlike traditional data terms for optical flow estimation that explain brightness changes in the image sequence exclusively by motion, our data term models changes in illumination as additional source for brightness variations. In order to estimate these changes jointly with the motion, we make use of a *parametrized* brightness transfer function (BTF) which was originally proposed by Grossberg and Nayar [17] in the context of photometric calibration for HDR imaging. This function maps intensities of the first frame to their intensities in the second frame. Given a set of n basis functions $\phi_j : \mathbb{R} \rightarrow \mathbb{R}$, the corresponding brightness transfer function reads

$$\Phi(\mathbf{c}, f) = \bar{\phi}(f) + \sum_{j=1}^n c_j \cdot \phi_j(f), \quad (2)$$

where $\bar{\phi} : \mathbb{R} \rightarrow \mathbb{R}$ is the mean brightness transfer function and $\mathbf{c} = (c_1, \dots, c_n)^\top$ are linear weights. Let us now discuss how to embed this general model for brightness changes into a data term. To this end, we propose the following combination of the *brightness* constancy assumption and the *gradient* constancy assumption which can be seen as an extension of [6]. Defining ν as positive weight and using $\nabla \cdot$ as spatial gradient operator, our data term reads

$$D(\mathbf{w}, \mathbf{c}) = D_{\text{bright}}(\mathbf{w}, \mathbf{c}) + \nu D_{\text{grad}}(\mathbf{w}, \mathbf{c}), \quad (3)$$

with

$$D_{\text{bright}}(\mathbf{w}, \mathbf{c}) = \Psi_d \left(\left(f_2(\mathbf{x} + \mathbf{w}) - \Phi(\mathbf{c}(\mathbf{x}), f_1(\mathbf{x})) \right)^2 \right), \quad (4)$$

and

$$D_{\text{grad}}(\mathbf{w}, \mathbf{c}) = \Psi_d \left(\left\| \nabla f_2(\mathbf{x} + \mathbf{w}) - \nabla \Phi(\mathbf{c}(\mathbf{x}), f_1(\mathbf{x})) \right\|_2^2 \right), \quad (5)$$

where brightness changes are now modeled to be spatially variant, i.e. with non-constant coefficients \mathbf{c} . Thus, we allow different brightness transfer functions Φ at each position. For both assumptions, the same sub-quadratic penalizer $\Psi_d(s^2) = 2\lambda_d^2(1 + s^2/\lambda_d^2)^{\frac{1}{2}}$ is used to render the approach more robust w.r.t. outliers [2].

Please note that flow variables and illumination coefficients have intentionally been distributed to different frames. This avoids products of unknowns when linearizing the assumptions later on and thus makes the minimization better tractable. Moreover, at first glance, it may seem counter-intuitive to combine our explicit estimation strategy with a gradient constancy assumption that is invariant under additive illumination changes. However, the additional gradient constancy term supports the estimation at those locations where the coefficients can not adapt or have not yet adapted perfectly to the illumination changes. This is for instance the case at the beginning of the estimation, when neither the flow nor the coefficients have converged to their final values yet.

2.2 Regularization Terms

Since the constancy assumptions in the data term may locally fail to provide any information, a spatial regularization of both the flow variables and the illumination coefficients is required. Moreover, it is not clear from the data term how to distribute observed brightness changes between motion and illumination. While the parametrization in terms of basis functions already provides a meaningful representation given by the coefficient fields, the concrete modeling of both regularization terms plays an important role in resolving this ambiguity. Let us now discuss how we model the two regularizers.

Flow Regularization. While first order regularization strategies have a long and successful tradition [22], recently, second-order smoothness terms received notable attention. In particular, such terms turned out to be highly useful for non-fronto-parallel motion, since they are tailored towards piecewise affine solutions [5,35,42,46]. Consequently, we make use of the following second-order regularizer that has already been used in the context of image denoising [26] and shape-from-shading [47]:

$$R_{\text{flow}}(\mathbf{w}) = \alpha \cdot \Psi_s \left(\|\mathcal{H}u\|_F^2 + \|\mathcal{H}v\|_F^2 \right). \quad (6)$$

Here, α is a positive weight, $\|\mathcal{H}\cdot\|_F$ is the Frobenius norm of the Hessian, and $\Psi_s(s^2) = 2\lambda_s^2(1+s^2/\lambda_s^2)^{\frac{1}{2}}$ is a sub-quadratic penalizer that encourages piecewise affine solutions.

Coefficient Regularization. In contrast to the flow regularizer that models a piecewise affine flow field, we assume that neighboring pixels are subject to similar illumination changes, i.e. that the coefficients of the basis functions are piecewise constant. Additionally, discontinuities in the coefficient fields are assumed to be aligned with edges in the input images (e.g. shadow edges) [29]. Consequently, we follow the idea of Zimmer *et al.* [54] and employ the following anisotropic complementary regularization term

$$R_{\text{illum}}(\mathbf{c}) = \beta \cdot \sum_{i=1}^2 \Psi_{\text{illum}}^i \left(\sum_{j=1}^n \gamma_j (\mathbf{r}_i^\top \nabla c_j)^2 \right), \quad (7)$$

where β is a positive weight and the two directions \mathbf{r}_1 and $\mathbf{r}_2 = \mathbf{r}_1^\perp$ allow to adapt the smoothing direction locally across and along image edges, respectively. As proposed in [54], these directions can be derived as the eigenvectors of the so-called regularization tensor. In our case, this tensor must be computed from the *photometric uncompensated* first frame f_1 to ensure that brightness information related to illumination changes is *not* discarded. Moreover, all coefficient fields are regularized jointly with a single penalizer function per direction, since spatial changes of the brightness transfer function typically result in discontinuities in all coefficient fields. In this context, the derivatives of the coefficients have to be balanced with weights γ_j to reflect the different magnitude ranges of the coefficient fields. How we can estimate these weights together with the basis functions is discussed in Section 4. Finally, we have to define the penalizer functions. As suggested in [54], we use the edge-enhancing Perona-Malik regularizer

$\Psi_{\text{illum}}^1(s^2) = \lambda_c^2 \log(1 + s^2/\lambda_c^2)$ as penalizer across edges (in \mathbf{r}_1 -direction), while we apply the edge-preserving Charbonnier regularizer $\Psi_{\text{illum}}^2(s^2) = 2\lambda_c^2(1 + s^2/\lambda_c^2)^{\frac{1}{2}}$ along them (in \mathbf{r}_2 -direction).

3 Minimization

In order to handle large displacements, we follow the warping strategy of Brox *et al.* [6]. We split the unknowns, i.e. the flow field \mathbf{w} and the coefficient fields \mathbf{c} , into a known part $\mathbf{w}^k, \mathbf{c}^k$ and an unknown increment $d\mathbf{w}^k, d\mathbf{c}^k$ and embed their overall estimation into a coarse-to-fine fixed point iteration. Moreover, within this (outer) fixed point iteration, we linearize the brightness and the gradient constancy assumptions in terms of the flow increments $d\mathbf{w}^k$ such that we finally approximate the original non-convex optimization problem by a series of convex optimization problems.

Actually, this strategy comes down to solving a differential formulation of the original energy (1) at each level k of the coarse-to-fine approach. If we denote the first frame by $f_1^k = f_1(x, y)$ and the motion compensated second frame by $f_2^k = f_2(x+u^k, y+v^k)$, the corresponding differential formulation of data and smoothness terms is given as follows. While the brightness and gradient constancy terms become

$$D_{\text{bright}}^k = \Psi_d \left(\theta \cdot \left(f_{2,x}^k du^k + f_{2,y}^k dv^k + f_2^k - \bar{\phi}(f_1^k) - \sum_{j=1}^n (c_j^k + dc_j^k) \cdot \phi_j(f_1^k) \right)^2 \right) \quad (8)$$

and

$$D_{\text{grad}}^k = \Psi_d \left(\left\| \begin{pmatrix} \theta_x & 0 \\ 0 & \theta_y \end{pmatrix} \left(\begin{aligned} & \nabla f_{2,x}^k du^k + \nabla f_{2,y}^k dv^k + \nabla f_2^k \\ & - \sum_{j=1}^n \phi_j(f_1^k) \cdot \nabla (c_j^k + dc_j^k) \\ & - \left(\bar{\phi}'(f_1^k) + \sum_{j=1}^n (c_j^k + dc_j^k) \cdot \phi_j'(f_1^k) \right) \cdot \nabla f_1^k \end{aligned} \right\|_2^2 \right), \quad (9)$$

respectively, the flow regularizer is given by

$$R_{\text{flow}}^k = \alpha \cdot \Psi_s \left(\|\mathcal{H}(u^k + du^k)\|_F^2 + \|\mathcal{H}(v^k + dv^k)\|_F^2 \right), \quad (10)$$

and the coefficient regularizer reads

$$R_{\text{illum}}^k = \beta \cdot \sum_{i=1}^2 \Psi_{\text{illum}}^i \left(\sum_{j=1}^n \gamma_j (\mathbf{r}_i^\top \nabla (c_j^k + dc_j^k))^2 \right). \quad (11)$$

Additionally, constraint normalization has been applied to all constancy assumptions in terms of the weights θ, θ_x and θ_y as proposed in [44] for linearized constraints with more than two variables.

After discretizing the Euler-Lagrange equations of this differential energy with finite differences, we obtain a nonlinear system of equations due to the derivatives of the subquadratic penalizers Ψ_* . This nonlinear system is then solved by another (inner) fixed point iteration: All nonlinear expressions are repeatedly kept fixed, and the resulting linear systems are solved using the successive overrelaxation (SOR) method [52].

In order to speed up computations, the inner fixed point iteration is embedded in a cascadic multigrid scheme, i.e. finer levels are initialized with coarse-scale solutions [4]. This also explains why we use an incremental computation of the coefficient fields c^k although the corresponding expressions are linear in the original functional (1). As the cascadic multigrid approach always starts the computation from scratch, we need an increment (dc^k) which can be initialized to zero without losing all previous information.

4 Basis Learning for Brightness Transfer Functions

In the previous sections, we have introduced our novel variational model and have sketched how to minimize the corresponding energy. The goal of this section is to explain how we estimate the mean BTF $\bar{\phi}(f)$, the basis functions ϕ_j as well as the associated weights γ_j for the joint regularizer of the coefficient fields. The basic strategy of this paper is inspired by the *Empirical Model of Response (EMoR)* of Grossberg and Nayar [17], where the camera response function of imaging systems is parametrized with a set of basis functions. However, our model acts on intensities instead of irradiances. As already mentioned in Section 2, input intensities f are mapped to output intensities via

$$\Phi(f) = \bar{\phi}(f) + \sum_{j=1}^n c_j \cdot \phi_j(f). \quad (12)$$

Note that many kinds of polynomial and exponential illumination models can be represented using the appropriate basis functions. For instance, the affine model of Negahdaripour [30] fits into this framework by choosing

$$\bar{\phi}(f) = 0, \quad \phi_1(f) = 1, \quad \phi_2(f) = f. \quad (13)$$

The same holds also for the purely additive models in [7,28], i.e. if instead $\phi_2(f) = 0$.

The recent KITTI Vision Benchmark Suite [14] offers a huge set of real-world image sequences together with ground truth optical flow fields. This gives us access to samples of input and output intensity levels of realistic scenarios. In particular, the availability of optical flow fields allows us to register consecutive frames and to analyze the behavior of the true BTF on a per-pixel basis.

Our general strategy to learn a basis from this massive amount of training data consists of three steps: In a first step, we segment and cluster the training images according to illumination changes along the ground truth flow. The segmentation is important, since we cannot expect that different image pairs provide fundamentally different *global* BTFs. Instead, we have to estimate multiple BTFs per image pair, since typical illumination changes such as drop shadows or specular reflections are *local* phenomena. In a second step, we use the segmented input images and compute for each region of each input image a separate BTF. In a third step, all these BTFs are used to perform a

principal component analysis (PCA) in order to identify the most representative basis functions for the observed illumination changes. It is worth noting that these steps can be applied iteratively, i.e. the estimated basis functions can be used again to segment the input images and thus to obtain improved BTFs. Let us now detail on these steps.

Segmenting Illumination Changes. Let us assume that we are given the training image sequences with corresponding ground truth flows. In order to discriminate the image regions with distinct lighting situations, i.e. with different brightness transfer functions, for each image pair, we first have to determine the pointwise BTF for every pixel. However, although this pointwise BTF can be arbitrarily complex, the given images provide *only one* constraint per pixel: The unknown BTF must map the intensity of this pixel in the first frame to the intensity of the corresponding pixel in the second frame.

To relax this extremely under-determined problem, let us now assume we are already given an estimate of the basis functions. Then, the sought pointwise BTF can be approximated using the given basis, and our task comes down to computing the optimal coefficient vector c in each pixel. Consequently, this problem fits perfectly into our variational model from Section 2, with the difference that we *only* have to solve for the coefficients c , since the optical flow is given and does not need to be estimated. However, as the ground truth might not be provided at every pixel (i.e. due to occlusions or due to sparse laser scans), we have to disable the data term at those positions where a flow vector is missing. Basically, this procedure leads to a variational inpainting method [49], because if the data term is disabled, only the coefficient regularization term contributes to the energy. Note that unlike in traditional inpainting scenarios, we are not interested in the coefficient values at positions with missing data. We only want to enforce global communication in order to avoid isolated estimates.

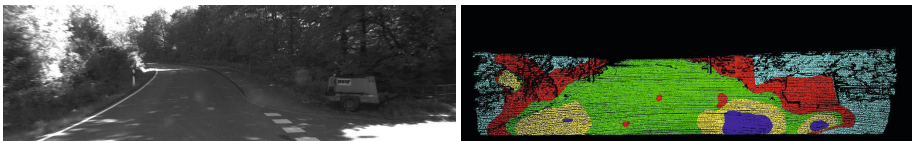


Fig. 1. **Left:** Frame 1 of KITTI training sequence #114. **Right:** Corresponding K-Means segmentation. Each color indicates a separate cluster, black pixels denote locations where ground truth is missing. The separation between the stronger brightening effect on the street and the weaker brightening effect in the environment becomes visible. Moreover, the inter-reflections at the windshield show off in terms of the three red spots on the street.

Once the coefficients are found, we perform a K -Means clustering (usually $K = 5$) on the coefficients. This takes place exclusively in the n -dimensional coefficient space; spatial coordinates are intentionally ignored here in order to allow spatially disjoint regions belonging to the same segment. All pixels whose coefficients have been clustered together share a similar brightness transfer function and thus exhibit a similar lighting situation. Figure 1 shows an example where such a segmentation allows to distinguish regions in the image that undergo different brightening effects.

Estimating Brightness Transfer Functions. Given the previously computed segmentation, the next task is to estimate one brightness transfer function $g : \mathbb{R} \rightarrow \mathbb{R}$ per segment. To this end, we adopt the *global* idea of Grossberg and Nayar [16] *locally*: For each segment we construct the intensity histogram h_1 of the pixels in the first frame as well as the histogram h_2 of the corresponding intensities in the second frame. In this context, we only consider pixels with valid optical flow, i.e. a ground truth vector must be given and must not point out of the image domain. Once both histograms have been created, we compute the BTF that transforms h_1 into h_2 by means of a histogram specification. In this context, fully saturated segments or too small clusters may lead to wrong and unrealistic brightness transfer functions. To avoid this, we reject any segments in which more than 80% of all pixels have the same intensity, as well as segments in which more than one third of all possible intensities do not occur. Please note that the resulting function of the histogram specification is discrete and given by an arbitrary vector $\mathbf{g} \in \mathbb{R}^{256}$ that is not parametrized in terms of basis functions and coefficients. Along with the BTFs from other segments it serves as input for the following PCA.

Learning the Basis. After having performed the previous clustering and estimation steps on each of the p training image pairs we obtain $m \leq K \cdot p$ brightness transfer functions, so-called *observations*. In order to find one common set of basis functions for all of them, we perform a principal component analysis (PCA). After concatenating all observations \mathbf{g}_i ($i = 1, \dots, m$) into a so-called *observation matrix*

$$\mathbf{G} = (\mathbf{g}_1 | \dots | \mathbf{g}_m) \in \mathbb{R}^{256 \times m}, \quad (14)$$

we compute the row-wise mean (i.e. the sample mean over all observations) $\bar{\mathbf{g}}$ of \mathbf{G} . Then we obtain the covariance matrix \mathbf{C} as

$$\mathbf{C} = \mathbf{U}^\top \mathbf{\Sigma} \mathbf{U} = \frac{1}{m-1} \sum_{i=1}^m (\mathbf{g}_i - \bar{\mathbf{g}})(\mathbf{g}_i - \bar{\mathbf{g}})^\top. \quad (15)$$

From this principal component decomposition, the sought basis functions ϕ_j ($j = 1, \dots, n$) can be found as the eigenvectors of the covariance matrix (the columns of \mathbf{U}). Moreover, the row-wise mean $\bar{\mathbf{g}}$ coincides with the 0-th basis function which is the mean brightness transfer function $\bar{\phi}$. Furthermore, the diagonal matrix $\mathbf{\Sigma}$ contains the eigenvalues which represent the variance of the given data along the principal components. This is a well-suited estimate for the relative magnitude of the coefficients. Hence, we set the weights γ_j in the anisotropic coefficient regularization term (7) to be the inverse square roots of the eigenvalues.

Figure 2 shows the estimated bases for the KITTI Vision Benchmark Suite and compares it to an affine basis and the *EMoR* basis provided by [17]. We can see that compared to the *EMoR* basis our basis functions for the KITTI benchmark rather model illumination changes in the upper part of the dynamic range. Moreover, the mean brightness transfer function is rather linear, since we do not estimate a camera response function as in [17] but a mapping between intensities (where identity is expected as average).

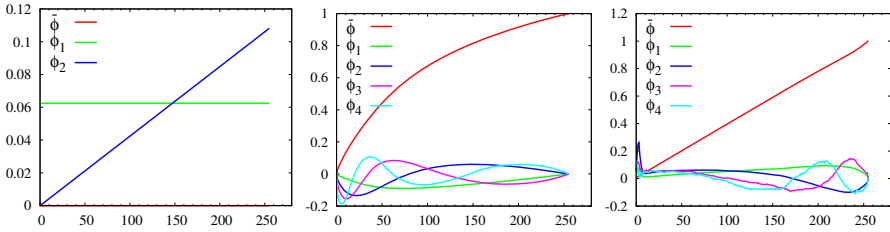


Fig. 2. Comparison of different basis functions. **From left to right: (a)** Normalized affine basis. **(b)** EMoR functions [16]. **(c)** Our basis functions learned from KITTI ground truth data.

Iterating the Estimation. The strategy we have described so far assumes a basis to be given for the clustering step. Initially, however, only the training images and ground truth flows are given. Thus, in our first iteration loop we omit the clustering step, treat the whole images as one segment, and estimate one global brightness transfer function per image pair. This leads to a first estimate for the basis which allows us then to perform the clustering as described. The impact of iterating the estimation of the basis functions on their shapes can be seen in Figure 3. While the mean BTF remains approximately the identity, the main support of the other basis functions is even further shifted towards the upper end of the dynamic range.

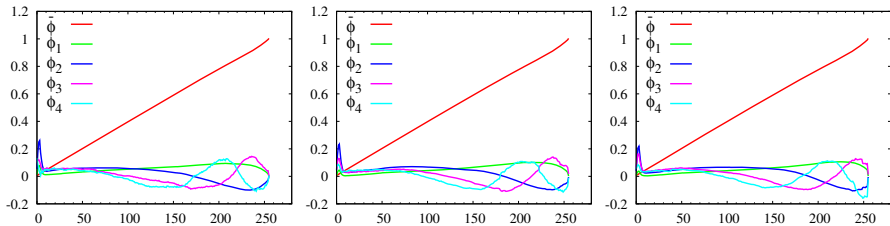


Fig. 3. Impact of iterating the estimation of the KITTI basis functions. **From left to right: (a)** Initial basis. **(b)** After one iteration. **(c)** After four iterations.

5 Evaluation

Our experiments are focused on the KITTI Vision Benchmark Suite [14] which offers a large amount of images depicting driving scenarios with challenging illumination changes. With our experiments we want to demonstrate that our method is very well suited for this kind of real-world imagery. The runtime of our single core implementation on an Intel XEON workstation with 3.2 GHz is about 80 seconds for each of the sequences (image size 1240×376). Results of our experiments are given in terms of the *bad pixel 3* (BP3) error measure that describes the percentage of estimated flow vectors that differ by more than 3 pixels from the respective ground truth.

Table 1. Comparison of different variants of our method on the full KITTI training set

Configuration	avg. BP3 error (occ)
Baseline (without illumination compensation)	11.17 %
Affine basis	11.07 %
EMoR basis	10.64 %
KITTI basis	10.71 %
KITTI basis (iterated)	10.19 %
KITTI basis (iterated, without gradient constancy)	10.65 %
KITTI basis (iterated, only gradient constancy)	10.95 %

Although our model contains a considerable number of parameters, effectively we set most of them fixed, and only adjusted the three main model parameters α , β , and ν . The contrast parameters for the sub-quadratic functions have been chosen fixed for all experiments as $(\lambda_d, \lambda_s, \lambda_{\text{illum}}^{1,2}) = (0.01, 0.5, 0.01)$. Concerning the K-Means clustering step we kept $K = 5$ fixed as well. For the number n of basis functions we found $n = 4$ to be a good tradeoff between computational effort and accuracy.

Evaluation of Basis Functions. In our first experiment, we investigate the usefulness of illumination estimation in general and analyze the impact of choosing different sets of basis functions on the quality of the flow estimation. To this end, we have computed the average BP3 errors of our method – with different bases as well as without any illumination compensation – on the provided set of 194 training sequences. For each configuration, we have optimized the parameters of our model w.r.t. the average BP3 error for the whole training set using the ground truth including occluded pixels (*occ*).

The results of this experiment are presented in Table 1. On the one hand, they clearly show that the benefit of using the classical affine model of Gennert and Negahdaripour [15] is rather limited compared to the baseline method without any illumination compensation – on average the error does hardly decrease. On the other hand, we can observe a clear improvement when choosing or learning a more suitable (and less ad-hoc) set of basis functions. Moreover, our experiments show that refining the basis functions iteratively allows to further improve the results. In the end, our proposed illumination model has been able to improve results for 81% of all training sequences. Finally, we also analyze the impact of the gradient constancy assumption. To this end, we consider variants of our method where the gradient constancy term has either been disabled or is the only constancy assumption. The two last rows of Table 1 show that in both cases the results deteriorate. This underlines our considerations from Section 2: The gradient term provides an improved initialization of the flow in early iterations, where the coefficient fields have not yet converged. However, the gradient term alone cannot provide sufficient information for estimating the basis coefficients since it considers only local differences, but establishes a mapping between absolute grey values.

Analysis of Transfer Functions and Coefficient Fields. In our second experiment we shed light on the coefficients that are estimated jointly with the optical flow. To this end, we have picked one of the training sequences with moderate illumination changes, see

Figure 4, and another sequence with severe illumination changes, see Figure 5. The two figures show the first frame of the respective sequence together with the estimated flow as well as the four computed coefficient fields. Furthermore, we have highlighted interesting locations in the images using colored squares. The brightness transfer functions at those locations – that can be computed as linear combinations of the learned basis functions weighted by the estimated local coefficients – are jointly depicted in a graph using the corresponding colors.

For our first challenging example (Figure 4), the flow field appears reasonably accurate, which is confirmed by a BP3 error of only 9.41%. Since the lighting changes in that image sequence are rather global, the extracted brightness transfer functions are similar in shape. In fact, they only differ in the upper end of the dynamic range. As can be seen from the BTFs, the image becomes darker. This is mainly reflected by the strongly negative values in the coefficient field c_1 (that belongs to a positive basis function). Moreover, slight local variations of the BTFs can be observed in the coefficient plots, in particular in the plots of the coefficient fields c_2, c_3, c_4 . An example, where our model has actually estimated significantly differing BTFs for different parts of the image is presented in Figure 5. Particularly challenging in this sequence are the inter-reflections in the windshield in front of the camera. However, the flow field is still of reasonable quality (BP3 of 5.33%). Apart of the spatially varying BTFs, one can also observe that the inter-reflections are reproduced by the corresponding coefficient fields.

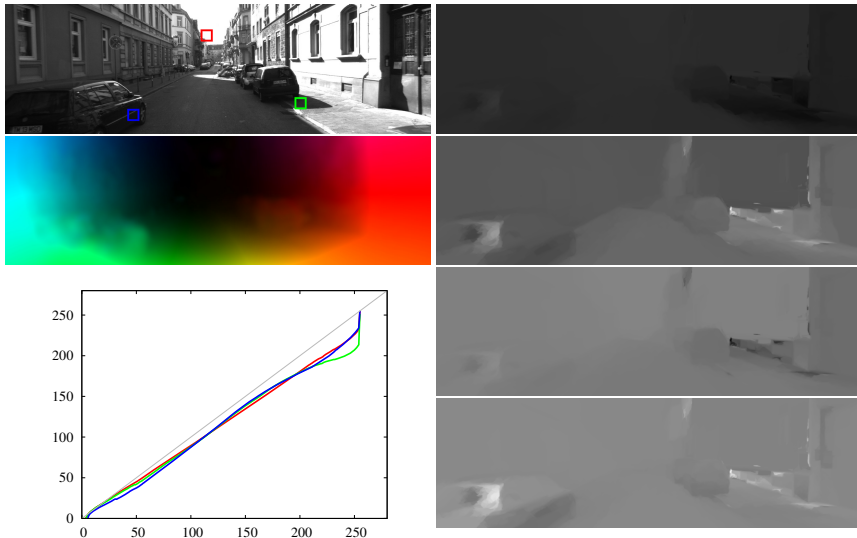


Fig. 4. Estimated coefficients and BTFs. **Left column, from top to bottom:** First frame of KITTI training sequence #15 with three highlighted positions, estimated optical flow field, plot of the three corresponding brightness transfer functions. Plot colors coincide with the marker colors. **Right column, from top to bottom:** Estimated coefficient fields c_1 to c_4 . Coefficients have been shifted such that a grey value of 127 denotes a coefficient of 0. Brighter values denote positive coefficients, darker values negative coefficients.

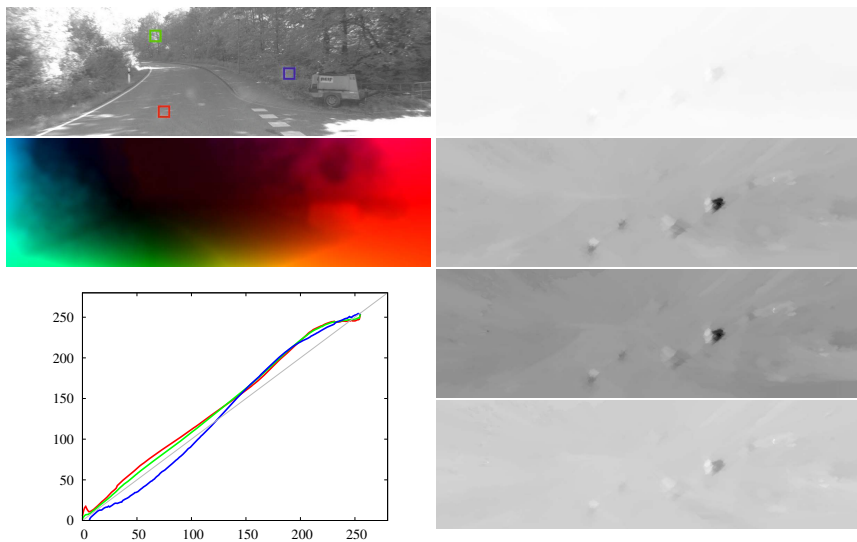


Fig. 5. Estimated coefficients and BTFs. **Left column, from top to bottom:** First frame of KITTI training sequence #114 with three highlighted positions, estimated optical flow field, plot of the three corresponding brightness transfer functions. Plot colors coincide with the marker colors. **Right column, from top to bottom:** Estimated coefficient fields c_1 to c_4 . Coefficients have been shifted such that a grey value of 127 denotes a coefficient of 0. Brighter values denote positive coefficients, darker values negative coefficients.

Table 2. Error statistics of our method for the bad pixel measure with varying thresholds (BP2 - BP5), averaged over all sequences of the KITTI evaluation benchmark

Error	Out-Noc	Out-All	Avg-Noc	Avg-All
2 pixels	8.84 %	14.14 %	1.5 px	2.8 px
3 pixels	6.52 %	11.03 %	1.5 px	2.8 px
4 pixels	5.38 %	9.29 %	1.5 px	2.8 px
5 pixels	4.64 %	8.11 %	1.5 px	2.8 px

Comparison to the Literature. In our third experiment, we compare our method to other approaches from the literature. To this end, we evaluated our method on the KITTI test sequences using the optimized parameters $(\alpha, \beta, \nu)=(5.2, 3, 6.25)$. The corresponding results are shown in Tables 2 and 3. While Table 2 gives detailed information on the performance of our algorithm in *non-occluded* and *all* regions for different thresholds of the bad pixel error measure (BP2 - BP5), Table 3 shows the performance of our algorithm compared to other *pure two-frame* optical flow methods *without stereo constraints* (such constraints are likely to fail in realistic scenarios with independently moving objects). As one can see, our method is among the leading optical flow approaches in this benchmark. In particular, when considering all pixels (i.e. also occluded regions), our method ranks first and is significantly more accurate than previous approaches. This clearly demonstrates that performing a joint estimation of illumination changes and motion can outperform methods discarding illumination information by using invariants.

Table 3. Comparison of pure two-frame optical flow methods for the KITTI evaluation sequences. Superscripts denote the rank of each method in the corresponding column at time of submission.

Method	Out-Noc	Out-All	Avg-Noc	Avg-All
DDR-DF	6.03 % ¹	13.08 % ²	1.6 px ⁵	4.2 px ³
TGV2ADCSIFT	6.20 % ²	15.15 % ⁴	1.5 px ²	4.5 px ⁴
Our method	6.52 %³	11.03 %¹	1.5 px²	2.8 px¹
Data-Flow	7.11 % ⁴	14.57 % ³	1.9 px ⁶	5.5 px ⁵
EpicFlow	7.19 % ⁵	16.15 % ⁵	1.4 px ¹	3.7 px ²
DeepFlow	7.22 % ⁶	17.79 % ⁶	1.5 px ²	5.8 px ⁷
TVL1-HOG	7.91 % ⁷	18.90 % ¹⁰	2.0 px ⁷	6.1 px ⁸
MLDP-OF	8.67 % ⁸	18.78 % ⁹	2.4 px ⁹	6.7 px ¹¹
DescFlow	8.76 % ⁹	19.45 % ¹¹	2.1 px ⁸	5.7 px ⁶
CRTflow	9.43 % ¹⁰	18.72 % ⁸	2.7 px ¹¹	6.5 px ⁹
C++	10.04 % ¹¹	20.26 % ¹²	2.6 px ¹⁰	7.1 px ¹²
C+NL	10.49 % ¹²	20.64 % ¹³	2.8 px ¹³	7.2 px ¹³
IVANN	10.68 % ¹³	21.09 % ¹⁴	2.7 px ¹¹	7.4 px ¹⁴
fSGM	10.74 % ¹⁴	22.66 % ¹⁵	3.2 px ¹⁵	12.2 px ¹⁵
TGV2CENSUS	11.03 % ¹⁵	18.37 % ⁷	2.9 px ¹⁴	6.6 px ¹⁰

6 Conclusions and Outlook

In this work we have addressed the problem of estimating the optical flow under uncontrolled illumination. In contrast to recent state-of-the-art methods that simply discard illumination information, we have proposed a novel variational model for jointly estimating both illumination changes and optical flow. In this context, we have contributed in three different ways: (i) In order to find a meaningful representation of illumination changes we have learned brightness transfer basis functions from previously segmented training data. (ii) By imposing a complementary regularizer on the corresponding coefficient fields we have been able to achieve a sharp separation between areas of different illumination changes while maintaining smoothness of the resulting flow field itself. (iii) By embedding both the basis functions and the coefficient regularization into a recent variational framework, we achieve state-of-the-art accuracy on the KITTI benchmark, outperforming competing approaches based on illumination-invariant assumptions.

This shows that approaches that additionally estimate relevant information, such as illumination changes, are a worthwhile alternative to approaches that simply discard that information for the sake of robustness. Moreover, such approaches can be used to provide subsequent algorithms with this additional information improving the overall performance. Future work includes the online learning of basis functions as well as the development of efficient numerical schemes on the GPU to speed up the computation.

References

1. Barnes, C., Shechtman, E., Goldman, D.B., Finkelstein, A.: The generalized PatchMatch correspondence algorithm. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part II. LNCS, vol. 6312, pp. 29–43. Springer, Heidelberg (2010)

2. Black, M.J., Anandan, P.: Robust dynamic motion estimation over time. In: Proc. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 292–302. IEEE Computer Society Press, Maui (1991)
3. Black, M.J., Fleet, D., Yacoob, Y.: Robustly estimating changes in image appearance. *Computer Vision and Image Understanding* 78(1), 8–31 (2000)
4. Bornemann, F., Deuffhard, P.: The cascadic multigrid method for elliptic problems. *Numerische Mathematik* 75, 135–152 (1996)
5. Braux-Zin, J., Dupont, R., Bartoli, A.: A general dense image matching framework combining direct and feature-based costs. In: Proc. IEEE International Conference on Computer Vision (ICCV), pp. 185–192. IEEE Press (2013)
6. Brox, T., Bruhn, A., Papenber, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: Pajdla, T., Matas, J.(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)
7. Cornelius, N., Kanade, T.: Adapting optical-flow to measure object motion in reflectance and X-ray image sequences. *Computer Graphics* 18(1), 24–25 (1984)
8. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Schmid, C., Soatto, S., Tomasi, C. (eds.) Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 886–893 (2005)
9. Debevec, P.E., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: Proc. SIGGRAPH 1997. Annual Conference Series, pp. 369–378. ACM Press (1997)
10. Dederscheck, D., Müller, T., Mester, R.: Illumination invariance for driving scene optical flow using comparagram preselection. In: Proc. IEEE Intelligent Vehicles Symposium (IV), pp. 742–747 (2012)
11. Demetz, O., Hafner, D., Weickert, J.: The complete rank transform: A tool for accurate and morphologically invariant matching of structures. In: Proceedings of the British Machine Vision Conference. BMVA Press (2013)
12. Garg, R., Roussos, A., Agapito, L.: A variational approach to video registration with subspace constraints. *International Journal of Computer Vision* 104(3), 286–314 (2013)
13. Garrido, P., Valgaerts, L., Wu, C., Theobalt, C.: Reconstructing detailed dynamic face geometry from monocular video. *ACM Transactions on Graphics* 32(6), 158:1–158:10 (2013)
14. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? The KITTI vision benchmark suite. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3354–3361. IEEE Computer Society (2012)
15. Gennert, M.A., Negahdaripour, S.: Relaxing the brightness constancy assumption in computing optical flow. Tech. Rep. 975, Artificial Intelligence Laboratory, Massachusetts Institute of Technology (June 1987)
16. Grossberg, M.D., Nayar, S.K.: What can be known about the radiometric response from images? In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part I. LNCS, vol. 2350, pp. 189–205. Springer, Heidelberg (2002)
17. Grossberg, M.D., Nayar, S.K.: Modeling the space of camera response functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(10), 1272–1282 (2004)
18. HaCohen, Y., Shechtman, E., Goldman, D.B., Lischinski, D.: Non-rigid dense correspondence with applications for image enhancement. *ACM Transactions on Graphics* 30(4), 70:1–70:9 (2011)
19. Hager, G.D., Belhumeur, P.N.: Real-time tracking of image regions with changes in geometry and illumination. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 403–410 (1996)
20. Haussecker, H.W., Fleet, D.J.: Estimating optical flow with physical models of brightness variation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(6), 661–673 (2001)

21. Hermosillo, G., Chefd'Hotel, C., Faugeras, O.: Variational methods for multimodal image matching. *International Journal of Computer Vision* 50(3), 329–343 (2002)
22. Horn, B., Schunck, B.: Determining optical flow. *Artificial Intelligence* 17, 185–203 (1981)
23. Kim, T.H., Lee, H.S., Lee, K.M.: Optical flow via locally adaptive fusion of complementary data costs. In: *Proc. IEEE International Conference on Computer Vision (ICCV)*, pp. 3344–3351. IEEE Press (2013)
24. Liu, C., Yuen, J., Torralba, A.: SIFT flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(5), 978–994 (2011)
25. Lowe, D.L.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
26. Lysaker, M., Lundervold, A., Tai, X.C.: Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *Transactions on Image Processing* 12(12), 1579–1590 (2003)
27. Mileva, Y., Bruhn, A., Weickert, J.: Illumination-robust variational optical flow with photometric invariants. In: Hamprecht, F.A., Schnörr, C., Jähne, B. (eds.) *DAGM 2007. LNCS*, vol. 4713, pp. 152–162. Springer, Heidelberg (2007)
28. Mukawa, N.: Estimation of shape, reflection coefficients and illuminant direction from image sequences. In: *Proc. IEEE International Conference on Computer Vision (ICCV)*, pp. 507–512 (1990)
29. Nagel, H.H., Enkelmann, W.: An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8, 565–593 (1986)
30. Negahdaripour, S., Yu, C.H.: A generalized brightness change model for computing optical flow. In: *Proc. IEEE International Conference on Computer Vision (ICCV)*, pp. 2–11. IEEE Computer Society (1993)
31. Nir, T., Bruckstein, A.M., Kimmel, R.: Over-parameterized variational optical flow. *International Journal of Computer Vision* 76(2), 205–216 (2008)
32. Ochs, P., Malik, J., Brox, T.: Segmentation of moving objects by long term video analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2013), early Access
33. Onkarappa, N., Sappa, A.: Speed and texture: an empirical study on optical-flow accuracy in ADAS scenarios. *IEEE Transactions on Intelligent Transportation Systems* 15(1), 136–147 (2014)
34. Papenberg, N., Bruhn, A., Brox, T., Didas, S., Weickert, J.: Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision* 67(2), 141–158 (2006)
35. Ranftl, R., Gehrig, S., Pock, T., Bischof, H.: Pushing the limits of stereo using variational stereo estimation. In: *IEEE Intelligent Vehicles Symposium*, pp. 401–407 (2012)
36. Rashwan, H.A., Mohamed, M.A., García, M.A., Mertsching, B., Puig, D.: Illumination robust optical flow model based on histogram of oriented gradients. In: Weickert, J., Hein, M., Schiele, B. (eds.) *GCPD 2013. LNCS*, vol. 8142, pp. 354–363. Springer, Heidelberg (2013)
37. Sadek, R., Facciolo, G., Arias, P., Caselles, V.: A variational model for gradient-based video editing. *International Journal of Computer Vision* 103(1), 127–162 (2013)
38. Schnörr, C.: On functionals with greyvalue-controlled smoothness terms for determining optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(10), 1074–1079 (1993)
39. Stein, F.J.: Efficient Computation of Optical Flow Using the Census Transform. In: Rasmussen, C.E., Bülthoff, H.H., Schölkopf, B., Giese, M.A. (eds.) *DAGM 2004. LNCS*, vol. 3175, pp. 79–86. Springer, Heidelberg (2004)

40. Steinbrücker, F., Pock, T., Cremers, D.: Advanced data terms for variational optic flow estimation. In: Magnor, M.A., Rosenhahn, B., Theisel, H. (eds.) *Proceedings of the Vision, Modeling, and Visualization Workshop (VMV)*, pp. 155–164. DNB (2009)
41. Sun, D., Roth, S., Black, M.J.: A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision* 106(2), 115–137 (2014)
42. Trobin, W., Pock, T., Cremers, D., Bischof, H.: An unbiased second-order prior for high-accuracy motion estimation. In: Rigoll, G. (ed.) *DAGM 2008*. LNCS, vol. 5096, pp. 396–405. Springer, Heidelberg (2008)
43. Uras, S., Girosi, F., Verri, A., Torre, V.: A computational approach to motion perception. *Biological Cybernetics* 60, 79–87 (1988)
44. Valgaerts, L., Bruhn, A., Zimmer, H., Weickert, J., Stoll, C., Theobalt, C.: Joint estimation of motion, structure and geometry from stereo sequences. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010, Part IV*. LNCS, vol. 6314, pp. 568–581. Springer, Heidelberg (2010)
45. van de Weijer, J., Gevers, T.: Robust optical flow from photometric invariants. In: *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 1835–1838 (2004)
46. Vogel, C., Roth, S., Schindler, K.: An evaluation of data costs for optical flow. In: Weickert, J., Hein, M., Schiele, B. (eds.) *GCPR 2013*. LNCS, vol. 8142, pp. 343–353. Springer, Heidelberg (2013)
47. Vogel, O., Bruhn, A., Weickert, J., Didas, S.: Direct shape-from-shading with adaptive higher order regularisation. In: Sgallari, F., Murli, A., Paragios, N. (eds.) *SSVM 2007*. LNCS, vol. 4485, pp. 871–882. Springer, Heidelberg (2007)
48. Wedel, A., Pock, T., Zach, C., Bischof, H., Cremers, D.: An improved algorithm for $TV-l^1$ optical flow. In: Cremers, D., Rosenhahn, B., Yuille, A.L., Schmidt, F.R. (eds.) *Statistical and Geometrical Approaches to Visual Motion Analysis*. LNCS, vol. 5604, pp. 23–45. Springer, Heidelberg (2009)
49. Weickert, J., Welk, M.: Tensor field interpolation with PDEs. In: Weickert, J., Hagen, H. (eds.) *Visualization and Processing of Tensor Fields*, pp. 315–325. Springer (2006)
50. Werlberger, M., Pock, T., Bischof, H.: Motion estimation with non-local total variation regularization. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2010, pp. 2464–2471 (2010)
51. Xu, L., Jia, J., Matsushita, Y.: Motion detail preserving optical flow estimation. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1293–1300. IEEE Computer Society Press (2010)
52. Young, D.M.: *Iterative Solution of Large Linear Systems*. Academic Press, New York (1971)
53. Zabih, R., Woodfill, J.: Non-parametric local transforms for computing visual correspondence. In: Eklundh, J.-O. (ed.) *ECCV 1994*. LNCS, vol. 800, pp. 151–158. Springer, Heidelberg (1994)
54. Zimmer, H., Bruhn, A., Weickert, J.: Optic flow in harmony. *International Journal of Computer Vision* 93(3), 368–388 (2011)