OPTICAL FLOW

FLORIAN BECKER, STEFANIA PETRA, CHRISTOPH SCHNÖRR

ABSTRACT. Motions of physical objects relative to a camera as observer naturally occur in everyday live and in many scientific applications. Optical flow represents the corresponding motion induced on the image plane. This paper describes the basic problems and concepts related to optical flow estimation together with mathematical models and computational approaches to solve them. Emphasis is placed on common and different modelling aspects and to relevant research directions from a broader perspective. The state of the art and corresponding deficiencies are reported along with directions of future research. The presentation aims at providing an accessible guide for practitioners as well as stimulating research work in relevant fields of mathematics and computer vision.

CONTENTS

1. INTRODUCTION

1.1. **Motivation, Overview.** *Motion of image data* belongs to the crucial features that enable low-level image analysis in natural vision systems, in machine vision systems, and the analysis of a major part of stored image data in the format of videos, as documented for instance by the fast increasing download rate of YouTube. Accordingly, *image motion analysis* has played a key role from the beginning of research in mathematical and computational approaches to image analysis.

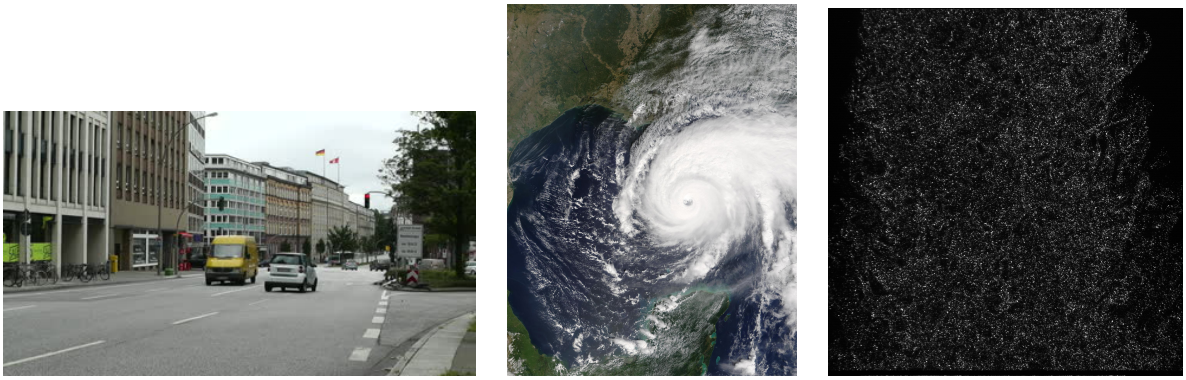


FIGURE 1. Some application areas of image processing that essentially rely on image motion analysis. LEFT: Scene analysis (depth, independently moving objects) with a camera mounted in a car. CENTER: Flow analysis in remote sensing. RIGHT: Measuring turbulent flows by particle image velocimetry.

Fig. 1 illustrates few application areas of image processing, among many others, where image motion analysis is deeply involved. Mathematical models for analyzing such image sequences boil down to models of a specific instance of the general data analysis task, that is to fuse prior knowledge with information given by observed image data. While adequate prior knowledge essentially depends on the application area as Fig. 1 indicates, the processing of observed data mainly involves basic principles that apply to *any* image sequence. Correspondingly, the notion of *optical flow*, informally defined as *determining the apparent instantaneous velocity of image structure*, emphasizes the application-independent aspects of this basic image analysis task.

Due to this independency, optical flow algorithms provide a key component for numerous approaches to *applications across different fields*. Major examples include motion compensation for video compression, structure from motion to estimate 3D scene layouts from an image sequences, visual odometry and incremental construction of mappings of the environment by autonomous systems, estimating vascular wall shear stress from blood flow image sequences for biomedical diagnosis, to name just a few.

This chapter aims at providing a concise and up-to-date account of mathematical models of optical flow estimation. Basic principles are presented along with various prior models. Application specific aspects are only taking into account at a general level of mathematical modeling (e.g., geometric or physical prior knowledge). Model properties favoring a particular direction of modeling are highlighted, while keeping an eye on common aspects and open problems. Conforming to the editor's guidelines, references to the literature are confined to a – subjectively defined – essential minimum.

1.2. **Organization.** Section 2 introduces a dichotomy of models used to present both essential differences and common aspects. These classes of models are presented in Sections 3 and 4. The former

class comprises those algorithms that perform best on current benchmark datasets. The latter class becomes increasingly more important in connection with motion analysis of novel, challenging classes of image sequences and videos. While both classes merely provide different viewpoints on the same subject – optical flow estimation and image motion analysis – distinguishing them facilitates the presentation of various facets of relevant mathematical models in current research. Further relationships, unifying aspects together with some major open problems and research directions, are addressed in Sect. 5.

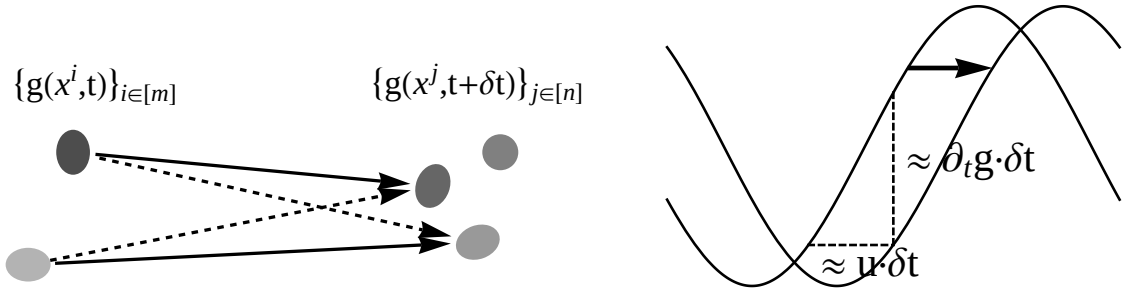


FIGURE 2. LEFT: Image motion can only be computed by recognizing objects as the same in subsequent time frames, based on some notion of equivalence (invariance) and some distance function. In low-level vision, "object" means some prominent local image structure in terms a feature mapping $g(x,t) \in \mathbb{R}^p$, $p \geq 1$. The correspondence problem amounts to compute a corresponding assignment $\{g(x^i,t)\}_{i\in[m]} \to \{g(x^j,t+\delta t)\}_{j\in[n]}$. The corresponding objective defines the data term of a variational approach. RIGHT: Differential approaches to image motion computation are based on smooth feature mappings $g(x,t)$ and aim at solving the assignment problem $g\big(x(t),t\big) \to g\big(x(t+\delta t),t+\delta t\big)$. The figure illustrates the basic case of a scalar-valued signal $g(x,t)$ translating with constant speed u and the estimate (2.14) based on the differential motion approach, as discussed in Sect. 2.2.

2. BASIC ASPECTS

2.1. **Invariance, Correspondence Problem.** Image motion computation amounts to define some notion of *invariance* and the recognition in subsequent time frames of *corresponding objects*, defined by local prominent image structure in term of a feature mapping $g(x)$ whose values are assumed to be conserved during motion. As Fig. 2, left panel, illustrates, invariance only holds approximately due to the imaging process and changes of viewpoint and illumination. Consequently, some *distance function*

$$\rho\big(g(x^j,t+\delta t) - g(x^i,t)\big) \tag{2.1}$$

has to be used in order to compute an *optimal assignment*

$$\{g(x^i,t)\}_{i\in[m]} \to \{g(x^j,t+\delta t)\}_{j\in[n]}. \tag{2.2}$$

A vast literature exists on definitions of feature mappings $g(x,t)$, distance functions, and their empirical evaluation in connection with image motion. Possible definitions include

- image grayvalue or color,
- grayvalue or color gradient,
- output of analytic bandpass filters (e.g. [86, 23]),
- more complex feature descriptors including SIFT [82] and SURF [17],
- censor voting, [27], local patches or feature groupings,

together with a corresponding invariance assumption, i.e. that $g(x,t)$ is conserved during motion (cf. Fig. 2, left panel). Figure 3 illustrates the most basic approaches used in the literature. Recent examples adopting a more geometric viewpoint on feature descriptors and studying statistical principles of patch similarity include [89, 105].

For further reference, some basic distance functions $\rho\colon \mathbb{R}^p \to \mathbb{R}_+$ are introduced below, that are commonly applied in connection with feature mappings $g(x)$ and partly parametrized by $\lambda > 0$ and

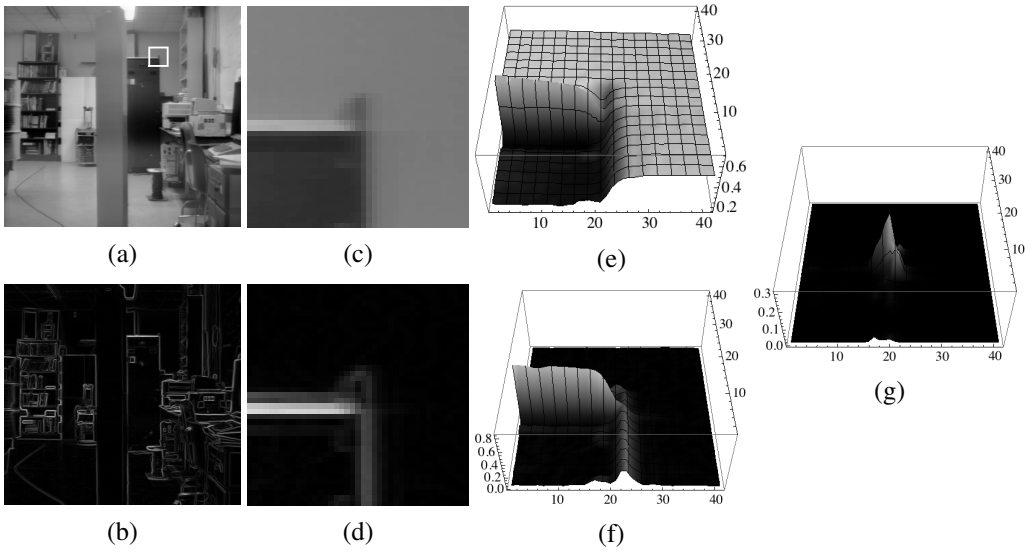


FIGURE 3. (a) Lab scene (©CMU image database) and (b) gradient magnitude that provides the basis for a range of feature mappings $g(x,t)$. The image section indicated in (a) is shown in (c), and (d) shows the same section extracted from (b). Panels (e) and (f) illustrate these sections as surface plots. Panel (g) shows a feature map responding to crossing grayvalue edges. (c), (d) and (g) correspond to the most basic examples of feature mappings $g(x,t)$ used in the literature to compute image motion, based on a corresponding invariance assumption (cf. Fig. 2, left panel) that is plausible for video frame rates.

$0 < \varepsilon \ll 1$. For closely related functions and the nomenclature in computer vision, see e.g. [24].

$$\rho_2^2(z) := \|z\|^2 \qquad \text{squared } \ell^2 \text{ distance,} \tag{2.3a}$$

$$\rho_2(z) := \|z\| \qquad \ell^2 \text{ distance,} \tag{2.3b}$$

$$\rho_{2,\varepsilon}(z) := \sqrt{\|z\|^2 + \varepsilon^2} - \varepsilon \qquad \text{smoothed } \ell^2 \text{ distance,} \tag{2.3c}$$

$$\rho_1(z) := \|z\|_1 \qquad \ell^1 \text{ distance,} \tag{2.3d}$$

$$\rho_{1,\varepsilon}(z) := \sum_{i \in [p]} \rho_{2,\varepsilon}(z_i) \qquad \text{smoothed } \ell^1 \text{ distance,} \tag{2.3e}$$

$$\rho_{2,\lambda}(z) := \min\{\|z\|^2, \lambda^2\} \qquad \text{truncated squared } \ell^2 \text{ distance,} \tag{2.3f}$$

$$\rho_{2,\lambda,\varepsilon}(z) := -\varepsilon \log \left(\mathrm{e}^{-\|z\|^2/\varepsilon} + \mathrm{e}^{-\lambda^2/\varepsilon} \right) \qquad \text{smoothed tr. sq. } \ell^2 \text{ distance.} \tag{2.3g}$$

Figure 4 illustrates these convex and non-convex distance functions. Functions $\rho_{1,\varepsilon}$ and $\rho_{2,\varepsilon}$ constitute specific instances of the general smoothing principle to replace a lower-semicontinuous, positively homogeneous and sublinear function $\rho(z)$ by a smooth proper convex function $\rho_\varepsilon(z)$, with $\lim_{\varepsilon \searrow 0} \varepsilon \rho_\varepsilon(z/\varepsilon) = \rho(z)$ (cf., e.g. [10]). Function $\rho_{2,\lambda,\varepsilon}(z)$ utilizes the log-exponential function [97, Ex. 1.30] to uniformly approximate $\rho_{2,\lambda}$ as $\varepsilon \searrow 0$.

2.2. **Assignment Approach, Differential Motion Approach.**

2.2.1. *Definitions.* Two basic approaches to image motion computation can be distinguished.

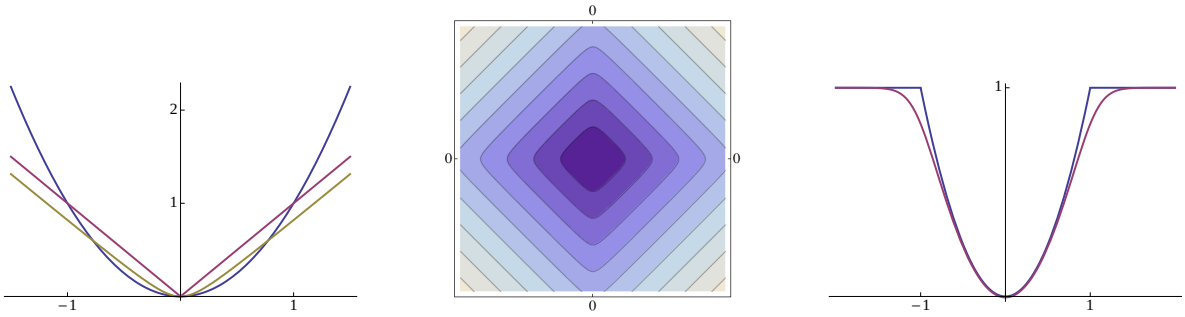


FIGURE 4. LEFT: Convex distance functions (2.3a)–(2.3c). CENTER: Level lines of the distance function $\rho_{1,\varepsilon}$ (2.3e). RIGHT: Non-convex distance functions (2.3f), (2.3g).

Assignment Approach, Assignment Field: This approach aims to determine an *assignment* of finite sets of spatially discrete features in subsequent frames of a given image sequence (Fig. 2, left panel). The vector field

$$u(x,t), \qquad x^j = x^i + u(x^i,t), \tag{2.4}$$

representing the assignment in Eq. (2.2), is called *assignment field*. This approach conforms to the basic fact that image sequences $f(x,t)$, $(x,t) \in \Omega \times [0,T]$ are recorded by *sampling* frames

$$\{f(x, k \cdot \delta t)\}_{k \in \mathbb{N}} \tag{2.5}$$

along the time axis.

Assignment approaches to image motion will be considered in Sect. 4.

Differential Motion Approach, Optical Flow: Starting point of this approach is the *invariance assumption* (Section 2.1) that observed values of some feature map $g(x,t)$ are conserved during motion,

$$\frac{\mathrm{d}}{\mathrm{d}t} g\big(x(t),t\big) = 0. \tag{2.6}$$

Evaluating this condition yields information about the trajectory $x(t)$ that represent the motion path of a particular feature value $g\big(x(t)\big)$. The corresponding vector field

$$\dot{x}(t) = \frac{\mathrm{d}}{\mathrm{d}t} x(t), \qquad x \in \Omega \tag{2.7}$$

is called *motion field* whose geometric origin will be described in Sect. 2.3. *Estimates*

$$u(x,t) \approx \dot{x}(t), \qquad x \in \Omega \tag{2.8}$$

of the motion field based on some observed time-dependent feature map $g(x,t)$, are called *optical flow fields*.

Differential motion approaches will be considered in Sect. 3.

2.2.2. *Common Aspects and Differences.* The assignment approach and the differential approach to image motion are closely related. In fact, for small temporal sampling intervals,

$$0 < \delta t \ll 1, \tag{2.9}$$

one may expect that the optical flow field multiplied by δt, $u(x,t) \cdot \delta t$, closely approximates the corresponding assignment field. The *same* symbol u is therefore used in (2.4) and (2.8) to denote the respective vector fields.

A conceptual difference between both approaches is that the ansatz (2.6) entails the assumption of a spatially differentiable feature mapping $g(x,t)$, whereas the assignment approach requires prior decisions done at a pre-processing stage that localize the feature sets (2.2) to be assigned. The need for additional processing in the latter case contrasts with the limited applicability of the differential approach: *The highest spatial frequency limits the speed of image motion $\|u\|$ that can be estimated reliably*:

$$\max\left\{\|\omega_x\|_\infty, \|u(x)\|\|\omega_x\| \colon \omega_x \in \operatorname{supp}\hat{g}(\omega),\ x \in \Omega\right\} \leq \frac{\pi}{6}. \tag{2.10}$$

The subsequent section details this bound in the most simple setting for a specific but common filter choice for estimating partial derivatives $\partial_i g$.

2.2.3. *Differential Motion Estimation: Case Study (1D).* Consider a scalar signal $g(x,t) = f(x,t)$ moving at constant speed (cf. Fig. 2, right panel),

$$\dot{x}(t) = \dot{x} = u, \qquad g\big(x(t),t\big) = g\big(x(0)+ut,t\big). \tag{2.11}$$

Note that the two-dimensional function $g(x,t)$ is a very special one generated by motion. Using the shorthands

$$x := x(0), \quad g_0(x) := g(x,0), \tag{2.12}$$

$g(x,t)$ corresponds to the translated one-dimensional signal

$$g(x,t) = g_0(x-ut) \tag{2.13}$$

due to the assumption $g\big(x(t),t\big) = g\big(x(0),0\big) = g_0(x)$.

Evaluating (2.6) at $t=0,\ x=x(0)$ yields

$$u = -\frac{\partial_t g_0(x)}{\partial_x g_0(x)} \quad \text{if} \quad \partial_x g_0(x) \neq 0. \tag{2.14}$$

Application and validity of this equation in practice depends on two further aspects: Only sampled values of $g(x,t)$ are given and the right-hand side has to be computed numericaly. Both aspects are discussed next in turn.

(1) In practice, samples are observed

$$\{g(k\cdot\delta x, t\delta t)\}_{k,t\in\mathbb{N}} = \{g(k,t)\}_{k,t\in\mathbb{N}}, \qquad \delta x = \delta t = 1, \tag{2.15}$$

with the sampling interval scaled to 1 without loss of generality. The Nyquist-Shannon sampling theorem imposes the constraint

$$\operatorname{supp}|\hat{g}(\omega)| \subset [0,\pi)^2, \qquad \omega = (\omega_x,\omega_t)^\top \tag{2.16}$$

where

$$\hat{g}(\omega) = \mathcal{F}g(\omega) = \int_{\mathbb{R}^2} g(x,t)e^{-i\langle\omega,\left(\begin{smallmatrix}x\\t\end{smallmatrix}\right)\rangle}dxdt \tag{2.17}$$

denotes the Fourier transform of $g(x,t)$. Trusting in the sensor, it may be savely assumed that $\operatorname{supp}|\hat{g}_0(\omega_x)| \subset [0,\pi)$. But what about the second coordinate t generated by motion? Does it obey (2.16) such that the observed samples (2.15) truly represent the one-dimensional video signal $g(x,t)$?

To answer this question, consider the specific case $g_0(x) = \sin(\omega_x x)$, $\omega_x \in [0,\pi]$ – see Fig. 5. Eq. (2.13) yields $g(x,t) = \sin\big(\omega_x(x-ut)\big)$. Condition (2.15) then requires that, for *every* location x, the one-dimensional time signal $g_x(t) := g(x,t)$ satisfies $\operatorname{supp}|\hat{g}_x(\omega_t)| \subset [0,\pi)$. Applying this to the example yields

$$g_x(t) = \sin(\omega_t t + \varphi_0), \qquad \omega_t := -\omega_x u, \quad \varphi_0 := \omega_x x, \tag{2.18}$$

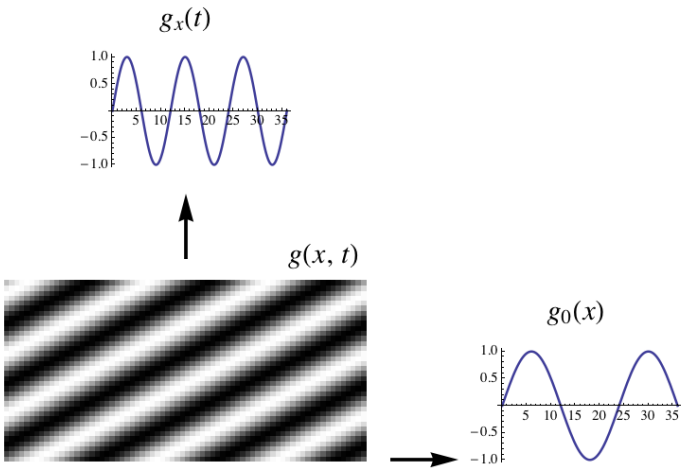


FIGURE 5. A sinusoid $g_0(x)$ with angular frequency $\omega_x = \pi/12$, translating with velocity $u = 2$, generates the function $g(x,t)$. The angular frequency of the signal $g_x(t)$ observed at a fixed position x equals $|\omega_t| = u \cdot \omega_x = \pi/6$ due to (2.18). It meets the upper bound further discussed in connection with Fig. 6 that enables accurate numerical computation of the partial derivatives of $g(x,t)$.

and hence the condition

$$|\omega_t| \in [0,\pi) \qquad \Leftrightarrow \qquad |u| < \frac{\pi}{\omega_x}. \tag{2.19}$$

It implies that equation (2.14) is only valid if, depending on the spatial frequency ω_x, the velocity u is sufficiently small.

This reasoning and the conclusion applies to general functions $g(x,t)$, $x \in \mathbb{R}^d$ in the form of (2.10), which additionally takes into account the effect of derivative estimation, discussed next.

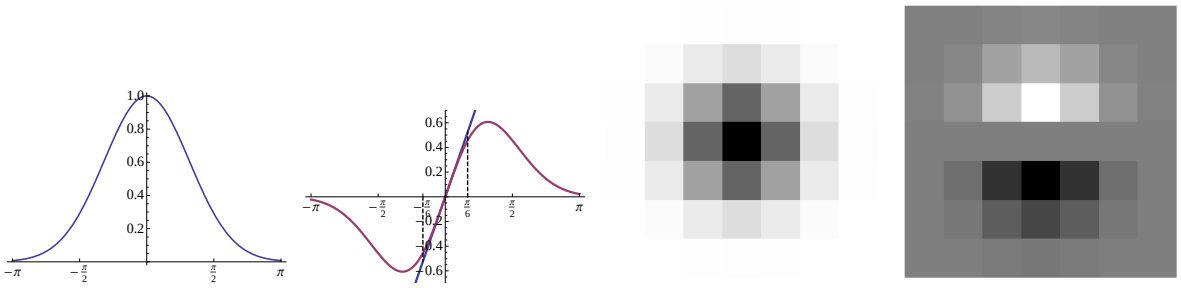
(2) Condition (2.19) has to be further restricted in practice, depending on how the partial derivatives of the r.h.s. of Eq. (2.14) are *numerically* computed using the observed samples (2.15). The Fourier transform

$$\mathcal{F}\big(\partial^\alpha g\big)(\omega) = i^{|\alpha|}\omega^\alpha \hat{g}(\omega), \qquad \omega \in \mathbb{R}^{d+1} \tag{2.20}$$

generally shows that taking partial derivatives of order $|\alpha|$ of $g(x,t)$, $x \in \mathbb{R}^d$, corresponds to high-pass filtering that amplifies noise. If $g(x,t)$ is vector-valued, then the present discussion applies to the computation of partial derivatives $\partial^\alpha g_i$ of any component $g_i(x,t)$, $i \in [p]$.

To limit the influence of noise, partial derivatives of the *low-pass filtered* feature mapping g are computed. This removes noise and smoothes the signal, and subsequent computation of partial derivatives becomes more accurate. Writing $g(x)$, $x \in \mathbb{R}^{d+1}$, instead of $g(x,t)$, $x \in \mathbb{R}^d$, to simplify the following formulas, low-pass filtering of g with the impulse response $h(x)$ means the convolution

$$g_h(x) := (h * g)(x) = \int_{\mathbb{R}^{d+1}} h(x-y)g(y)\mathrm{d}y, \qquad \hat{g}_h(\omega) = \hat{h}(\omega)\,\hat{g}(\omega) \tag{2.21}$$



(a) (b) (c) (d)

FIGURE 6. **(a)** Fourier transform $\hat{h}_\sigma(\omega)$ of the Gaussian low-pass (2.24), $\sigma = 1$. For values $\sigma \geq 1$, it satisfies the sampling condition $\operatorname{supp} |\hat{h}_\sigma(\omega)| \subset [0, \pi)$ sufficiently accurate. **(b)** The Fourier transform of the *Derivative-of-Gaussian (DoG)* filter $\frac{\mathrm{d}}{\mathrm{d}x} h_\sigma(x)$ illustrates that for $|\omega| \leq \pi/6$ (partial) derivatives are accurately computed while noise is suppressed at higher angular frequencies. **(c)**, **(d)** The impulse responses $h_\sigma(x,t)$ and $\partial_t h_\sigma(x,t)$ up to size $|x|, |t| \leq 2$. Application of the latter filter together with $\partial_x h_\sigma(x,t)$ to the function $g(x,t)$ discussed in connection with Fig. 5 and evaluation of Eq. (2.14) yield the estimate $u = 2.02469$ at all locations (x,t) where $\partial_x g(x,t) \neq 0$.

whose Fourier transform corresponds to the multiplication of the respective Fourier transforms. Applying (2.20) yields

$$\mathcal{F}\big(\partial^\alpha g_h\big)(\omega) = i^{|\alpha|}\omega^\alpha\big(\hat{h}(\omega)\,\hat{g}(\omega)\big) = \big(i^{|\alpha|}\omega^\alpha \hat{h}(\omega)\big)\hat{g}(\omega). \tag{2.22}$$

Thus, computing the partial derivative of the filtered function g_h can be computed by convolving g with the partial derivative of the impulse response $\partial^\alpha h$. As a result, the approximation of the partial derivative of g reads

$$\partial^\alpha g(x) \approx \partial^\alpha g_h(x) = \big((\partial^\alpha h) * g\big)(x). \tag{2.23}$$

The most common choice of h is the isotropic Gaussian low-pass filter

$$h_\sigma(x) := \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\Big(-\frac{\|x\|^2}{2\sigma^2}\Big) = \prod_{i\in[d]} h_\sigma(x_i), \qquad \sigma > 0. \tag{2.24}$$

that factorizes (called *separable filter*) and therefore can be implemented efficiently. The corresponding filters $\partial^\alpha h_\sigma(x)$, $|\alpha| \geq 1$, are called *Derivative-of-Gaussian (DoG) filters*.

To examine its effect, it suffices to consider any coordinate due to factorization, that is the one-dimensional case. Fig. 6 illustrates that values $\sigma \geq 1$ lead to filters that are sufficiently band-limited so as to conform to the sampling theorem. The price to pay for effective noise suppression however is a more restricted range $\operatorname{supp} |\mathcal{F}\big(g(x,t)\big)| = [0, \omega_{x,\max}]$, $\omega_{x,\max} \ll \pi$, that observed image sequence functions have to satisfy, so as to enable accurate computation of partial derivatives, and in turn accurate motion estimates based on the differential approach. Figure 5 further details and illustrates this crucial fact.

2.2.4. *Assignment or Differential Approach?* For image sequence functions $g(x,t)$ satisfying the assumptions necessary to evaluate the key equation (2.6), the differential motion approach is more convenient. Accordingly, much work has been devoted to this line of research up to now. In particular,

sophisticated multiscale representations of $g(x,t)$ enable to estimate larger velocities of image motion using smoothed feature mapping g (cf. Sect. 3.3.3). As a consequence, differential approaches rank top at corresponding benchmark evalutions conforming to the underlying assumptions [115] and efficient implementations are feasible [29, 55].

On the other hand, the inherent limitations of the differential approach discussed above become increasingly more important in current applications, like optical flow computation for traffic scenes taken from a moving car at high speed. Figure 1, right panel, shows another challenging scenario where the spectral properties $\hat{g}(\omega_x, \omega_t)$ of the image sequence function and the velocity fields to be estimated render application of the differential approach difficult, if not impossible. In such cases, the assignment approach is the method of choice.

Combining both approaches in a complementary way seems most promising: robust assignments enable to cope with fast image motions, and a differential approach turns these estimates into spatially dense vector fields. This point is taken up in Sect. 5.1.

2.2.5. *Basic Difficulties of Motion Estimation.* This section concludes with a list of some basic aspects to be addressed by any approach to image motion computation:

(i) Definition of a feature mapping g assumed to be conserved during motion (Sect. 2.1).
(ii) Coping with lack of invariance of g, change of appearance due to varying viewpoint and illumination (Sect. 3.3.1, 3.3.2).
(iii) Spatial sparsity of distinctive features (Sect. 3.4).
(iv) Coping with ambiguity of locally optimal feature matches (Sect. 4.2).
(v) Occlusion and disocclusion of features.
(vi) Consistent integration of available prior knowledge, regularization of motion field estimation (Sect. 3.5.2, 3.5.3).
(vii) Runtime requirements (Sect. 3.6).

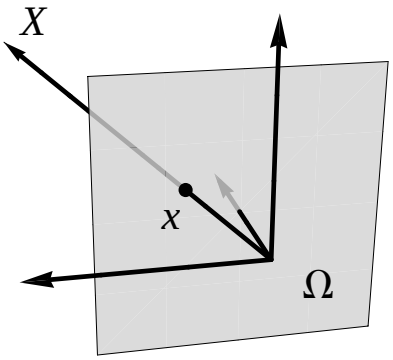


FIGURE 7. The basic pinhole model of the mathematically ideal camera. Scene points X are mapped to image points x by perspective projection.

2.3. **Two-View Geometry, Assignment and Motion Fields.** This section collects few basic relationships related to the Euclidean motion of a perspective camera relative to a 3D scene, that induces both the assignment field and the motion field on the image plane, as defined in Sect. 2.2.1 by (2.4) and

(2.7). Figures 7 and 12 illustrates these relationships. References [57, 45] provide comprehensive expositions.

It is pointed out once more that assignment and motion fields are purely *geometrical* concepts. The explicit expressions (2.43) and (2.53b) illustrate how discontinuities of these fields correspond to *discontinuities of depth*, or to *motion boundaries* that separate regions in the image plane of scene objects (or the background) with different motions relative to the observing camera. Estimates of either field will be called *optical flow*, to be discussed in subsequent sections.

2.3.1. *Two-View Geometry.* Scene and corresponding image points are denoted by $X \in \mathbb{R}^3$ and $x \in \mathbb{R}^2$, respectively. Both are incident with the line λx, $\lambda \in \mathbb{R}$, through the origin. Such lines are points of the projective plane denoted by $y \in \mathbb{P}^2$. The components of y are called *homogeneous coordinates* of the image point x, whereas x and X are the *inhomogeneous coordinates* of image and scene points, respectively. Note that y stands for *any* representative point on the ray connecting x and X. In other words, when using homogeneous coordinates, scale factors do not matter. This equivalence is denoted by

$$y \simeq y' \quad \Leftrightarrow \quad y = \lambda y', \quad \lambda \neq 0. \tag{2.25}$$

Figure 7 depicts the mathematical model of a pinhole camera with the image plane located at $X_3 = 1$. Perspective projection corresponding to this model connects homogeneous and inhomogeneous coordinates by

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{y_3} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}. \tag{2.26}$$

A *particular representative* y with *unknown depth* $y_3 = X_3$ equals the scene point X. This reflects the fact that scale cannot be inferred from a single image. The 3D space $\mathbb{R}^3 \setminus \{0\}$ corresponds to the affine chart $\{y \in \mathbb{P}^2 \colon y_3 \neq 0\}$ of the manifold $\mathbb{P}^2$.

Similar to representing an image point x by homogeneous coordinates y, it is common to represent scene points $X \in \mathbb{R}^3$ by *homogeneous coordinates* $Y = (Y_1, Y_2, Y_3, Y_4)^\top \in \mathbb{P}^3$, in order to linearize transformations of 3D space. The connection analogous to (2.26) is

$$X = \frac{1}{Y_4} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}. \tag{2.27}$$

Rigid (Euclidean) transformations are denoted by $\{h, R\} \in \mathsf{SE}(3)$ with translation vector h and proper rotation matrix $R \in \mathsf{SO}(3)$ characterized by $R^\top R = I$, $\det R = +1$. Application of the transformation to a scene point X and some representative Y reads

$$RX + h \qquad \text{and} \qquad QY, \qquad Q := \begin{pmatrix} R & h \\ 0^\top & 1 \end{pmatrix}, \tag{2.28}$$

whereas the inverse transformation $\{-R^\top h, R^\top\}$ yields

$$R^\top (X - h) \qquad \text{and} \qquad Q^{-1} Y, \qquad Q^{-1} = \begin{pmatrix} R^\top & -R^\top h \\ 0^\top & 1 \end{pmatrix}. \tag{2.29}$$

The nonlinear operation (2.26), entirely rewritten with homogeneous coordinates, takes the linear form

$$y = PY, \qquad P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = (I_{3\times 3}, 0), \tag{2.30}$$

with the *projection matrix* P and *external* or *motion parameters* $\{h, R\}$. In practice, additional *internal parameters* characterizing real cameras to the first order of approximation are taken into account in terms of a *camera matrix* K and the corresponding modification of (2.30),

$$y = PY, \qquad P = K(I_{3\times 3}, 0). \tag{2.31}$$

As a consequence, the transition to *normalized (calibrated)* coordinates

$$\tilde{y} := K^{-1}y \tag{2.32}$$

corresponds to an *affine* transformation of the image plane.

Given an image point x, taken with a camera in the canonical position (2.30), the corresponding ray meets the scene point X, see Figure 12 (b). This ray projects in a second image, taken with a second camera positioned by $\{h, R\}$ relative to the first camera and with projection matrix

$$P' = K'R^\top(I, -h), \tag{2.33}$$

to the line l', on which the projection x' of X *corresponding* to x must lie. Turning to homogeneous coordinates, an elementary computation shows that the *fundamental matrix*

$$F := K'^{-\top}R^\top[h]_\times K^{-1} \tag{2.34}$$

maps y to the *epipolar line* l',

$$l' = Fy. \tag{2.35}$$

This relation is symmetrical in that $F^\top$ maps y' to the corresponding epipolar line l in the first image,

$$l = F^\top y'. \tag{2.36}$$

The *epipoles* e, e' are the image points corresponding to the projection centers. Because they lie on l and l' for *any* x' and x, respectively, it follows that

$$Fe = 0, \qquad F^\top e' = 0. \tag{2.37}$$

The incidence relation $x' \in l'$ algebraically reads $\langle l', y'\rangle = 0$. Hence by (2.35),

$$\langle y', Fy\rangle = 0 \tag{2.38}$$

This key relation constrains the correspondence problem $x \leftrightarrow x'$ for arbitrary two views of the same unknown scene point X. Rewriting (2.38) in terms of *normalized* coordinates by means of (2.32) yields

$$\langle y', Fy\rangle = \langle K'^{-1}y', K'^{\top}FK(K^{-1}y)\rangle = \langle K'^{-1}y', E(K^{-1}y)\rangle \tag{2.39}$$

with the *essential matrix* E that, due to (2.34) and the relation $[Kh]_\times \simeq K^{-\top}[h]_\times K^{-1}$, is given by

$$E = K'^{\top}FK = R^\top[h]_\times. \tag{2.40}$$

Thus, essential matrices are parametrized by transformations $\{h, R\} \in \mathsf{SE}(3)$ and therefore form a smooth manifold embedded in $\mathbb{R}^{3\times 3}$.

2.3.2. *Assignment Fields.* Throughout this section, the internal camera parameters K are assumed to be known and hence normalized coordinates (2.32) are used. As a consequence,

$$K = I \tag{2.41}$$

is set in what follows.

Suppose some motion h, R of a camera relative to a 3D scene causes the image point x of a fixed scene point X to move to x' in the image plane. The corresponding *assignment vector* $u(x)$ represents the *displacement* of x in the image plane,

$$x' = x + u(x), \tag{2.42}$$

which due to (2.29) and (2.26) is given by

$$u(x) = \frac{1}{\langle r^3, X - h\rangle} \begin{pmatrix} \langle r^1, X - h\rangle \\ \langle r^2, X - h\rangle \end{pmatrix} - \frac{1}{X_3} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}. \tag{2.43}$$

Consider the special case of pure translation, i.e. $R = I$, $r^i = e^i$, $i = 1, 2, 3$. Then

$$u(x) = \frac{1}{X_3 - h_3} \begin{pmatrix} X_1 - h_1 \\ X_2 - h_2 \end{pmatrix} - \frac{1}{X_3} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \tag{2.44a}$$

$$= \frac{1}{\tilde{h}_3 - 1} \left(\begin{pmatrix} \tilde{h}_1 \\ \tilde{h}_2 \end{pmatrix} - \tilde{h}_3 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right), \qquad \tilde{h} := \frac{1}{X_3} h. \tag{2.44b}$$

The image point x_e where the vector field u vanishes, $u(x_e) = 0$, is called *focus of expansion (FOE)*

$$x_e = \frac{1}{\tilde{h}_3} \begin{pmatrix} \tilde{h}_1 \\ \tilde{h}_2 \end{pmatrix}. \tag{2.45}$$

x_e corresponds to the epipole $y = e$ since $Fe \simeq R^\top [h]_\times h = 0$.

Next the transformation is computed of the image plane induced by the motion of the camera in terms of projection matrices $P = (I, 0)$ and $P' = R^\top (I, -h)$ relative to a plane in 3D space

$$\langle n, X\rangle - d = n_1 X_1 + n_2 X_2 + n_3 X_3 - d = 0, \tag{2.46}$$

with unit normal n, $\|n\| = 1$, and with signed distance d of the plane from the origin 0. Setting $p = \left(\begin{smallmatrix} n \\ -d \end{smallmatrix}\right)$, Eq. (2.46) reads

$$\langle p, Y\rangle = 0. \tag{2.47}$$

In order to compute the point X on the plane satisfying (2.46) that projects to the image point y, the ray $Y(\lambda) = \left(\begin{smallmatrix} \lambda y \\ 1 \end{smallmatrix}\right)$, $\lambda \in \mathbb{R}$, is intersected with the plane.

$$\langle p, Y(\lambda)\rangle = \lambda \langle n, y\rangle - d = 0 \;\Rightarrow\; \lambda = \frac{d}{\langle n, y\rangle},\; Y = \begin{pmatrix} \frac{d}{\langle n, y\rangle} y \\ 1 \end{pmatrix} \simeq \begin{pmatrix} y \\ \frac{1}{d}\langle n, y\rangle \end{pmatrix}. \tag{2.48}$$

Projecting this point onto the second image plane yields

$$\begin{aligned} y' = P' Y(\lambda) &= R^\top \left(y - \frac{1}{d} \langle n, y\rangle h \right) = R^\top (I - \frac{1}{d} h n^\top) y \\ &=: Hy \end{aligned} \tag{2.49}$$

Thus, moving a camera relative to a 3D plane induces a *homography* (projective transformation) H of $\mathbb{P}^2$ which by virtue of (2.26) yields an assignment field $u(x)$ with *rational* components.

2.3.3. *Motion Fields.* Motion fields (2.7) are the instantaneous (differential) version of assignment fields. Consider a smooth path $\{h(t), R(t)\} \subset \mathsf{SE}(3)$ through the identity $\{0, I\}$ and the corresponding path of a scene point $X \in \mathbb{R}^3$

$$X(t) = h(t) + R(t) X, \qquad X = X(0). \tag{2.50}$$

Let $R(t)$ be given by a rotational axis $q \in \mathbb{R}^3$ and a rotation angle $\varphi(t)$. Using Rodrigues' formula and the skew-symmetric matrix $[q]_\times \in \mathfrak{so}(3)$ with $\dot\varphi = \dot\varphi(0) := \|q\|$, matrix $R(t)$ takes the form

$$R(t) = \exp(t[q]_\times) = I + \frac{\sin(\dot\varphi t)}{\dot\varphi t} t[q]_\times + \frac{1 - \cos(\dot\varphi t)}{(\dot\varphi t)^2} t^2 [q]_\times^2. \tag{2.51}$$

(2.50) then yields

$$\dot X(0) = v + [q]_\times X, \qquad v := \dot h(0), \tag{2.52}$$

where v is the translational velocity at $t = 0$. Differentiating (2.26) with $y = X$ (recall assumption (2.41)) and inserting (2.52), gives

$$\frac{\mathrm{d}}{\mathrm{d}t}x = \frac{1}{X_3^2}\begin{pmatrix} X_3\dot{X}_1 - X_1\dot{X}_3 \\ X_3\dot{X}_2 - X_2\dot{X}_3 \end{pmatrix} = \frac{1}{X_3}\begin{pmatrix} \dot{X}_1 - x_1\dot{X}_3 \\ \dot{X}_2 - x_2\dot{X}_3 \end{pmatrix} \tag{2.53a}$$

$$= \frac{1}{X_3}\left[\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} - v_3\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right] + \begin{pmatrix} q_2 - q_3x_2 - q_1x_1x_2 + q_2x_1^2 \\ -q_1 + q_3x_1 + q_2x_1x_2 - q_1x_2^2 \end{pmatrix}. \tag{2.53b}$$

Comparing (2.53b) to (2.43) and (2.44b) shows a similar structure of the translational part with FOE

$$x_v := \frac{1}{v_3}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \tag{2.54}$$

whereas the rotational part merely contributes an incomplete second-order degree polynomial to each component of the motion field, that do not depend on the scene structure in terms of the depth X_3.

Consider the special case of a motion field induced by the relative motion of a camera to a 3D plane given by (2.46) and write

$$\frac{1}{X_3} = \frac{1}{d}\left(n_3 + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix}^\top \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right). \tag{2.55}$$

Insertion into (2.53b) shows that the overall expression for the motion fields takes a simple polynomial form.

2.4. **Early Pioneering Work.** It deems proper to the authors to refer at least briefly to early pioneering work related to optical flow estimation, as part of a survey paper. The following references constitute just a small sample of the rich literature.

The information of motions fields, induced by the movement of an observer relative to a 3D scene, was picked out as a central theme more than three decades ago [81, 95]. Kanatani [72] studied the representation of $\mathsf{SO}(3)$ and invariants in connection with the space of motion fields induced by the movement relative to a 3D plane. Approaches to estimating motion fields followed soon, by determining optical flow from local image structure [84, 91, 63, 133, 131, 132]. Poggio and Verri [122] pointed out both the inexpedient, restrictive assumptions making the invariance assumption (2.6) hold in the simple case $g(x) = f(x)$ (e.g. Lambertian surfaces in the 3D scene), and the stability of structural (topological) properties of motion fields (like e.g. the FOE (2.45)). The local detection of image translation as orientation in spatio-temporal frequency space, based on the energy and the phase of collections of orientation-selective complex-valued bandpass filters (lowpass filters shifted in Fourier space, like e.g. Gabor filters), was addressed by [2, 60, 46], partially motivated by related research on natural vision systems.

The variational approach to optical flow was pioneered by Horn and Schunck [68], followed by various extensions [92, 6, 144] including more mathematically oriented accounts [109, 64, 130]. The work [129] classified various convex variational approaches that have unique unique minimizers.

The computation of discontinuous optical flow fields, in terms of piecewise parametric representations, was considered by [128, 24], whereas the work [118] studied the information contained in correspondences induced by motion fields over a longer time period. Shape-based optimal control of flows determined on discontinuous domains as control variable, was introduced in [110], including the application of shape derivative calculus that became popular later on in connection with level sets. Markov random fields and the Bayesian viewpoint on the non-local inference of discontinuous optical flow fields were introduced in [62]. The challenging aspects of estimating both motion fields and their segmentation in a spatio-temporal framework, together with inferring the 3D structure, has remained a topic of research until today.

This brief account shows that most of the important ideas appeared early in the literature. On the other hand, it took many years until first algorithms made their way into industrial applications. A lot of work remains to be done by addressing various basic and applied research aspects. In comparison to the fields of computer vision, computer science and engineering, not much work has been done by the mathematical community on motion based image sequence analysis.

2.5. **Benchmarks.** Starting with the first systematic evaluation in 1994 by Baron et al. [15], benchmarks for optical flow methods have stimulated and steered the developement of new algorithms in this field. The *Middlebury database* [14] further accelerated this trend by introducing an online ranking system and defining challenging data sets, which specifically address different aspects of flow estimation such as large displacements or occlusion.

The recently introduced *KITTI Vision Benchmark Suite* [47] concentrates on *outdoor* automotive sequences that are affected by disturbances such as illumination changes and reflections, which optical flow approaches are expected to be robust against.

While real imagery requires sophisticated measurement equipment to capture reliable reference information, *synthetic* sequences such as the novel *MPI Sintel Flow Dataset* [32] come with free ground truth. However, enormous efforts are necessary to *realistically* model the scene complexity and effects found in reality.

3. The Variational Approach to Optical Flow Estimation

In contrast to assignment methods, variational approaches to estimating the optical flow employ a *continuous* and *dense* representation of the variables $u : \Omega \mapsto \mathbb{R}^2$. The model describing the agreement of u with the image data defines the *data term* $E_{\mathrm{D}}(u)$. It is complemented by a *regularization term* $E_{\mathrm{R}}(u)$ encoding prior knowledge about the *spatial smoothness* of the flow. Together these terms define the *energy function* $E(u)$ and estimating the optical flow amounts to finding a global minimum u, possibly constrained by a set U of admissible flow fields, and using an appropriate numerical method:

$$\inf_{u \in U} E(u)\,, \quad E(u) := E_{\mathrm{D}}(u) + E_{\mathrm{R}}(u) \tag{3.1}$$

$E(u)$ is non-convex in general and hence only suboptimal solutions can be determined in practice.

Based on the variational approach published in 1981 by Horn & Schunck [68] a vast number of refinements and extensions were proposed in literature. Recent comprehensive empirical evaluations [47, 14] reveal that algorithms of this family yield best performance. Section 3.2 introduces the approach of Horn and Schunck as reference for the following discussion, after deriving the required linearized invariance assumption in Sect. 3.1.

Data and regularization terms designed to cope with various difficulties in real applications are presented in Sections 3.3 and 3.4, respectively. Section 3.6 gives a short overview over numerical algorithms for solving problem (3.1). Section 3.5 addresses some important extensions of the discussed framework.

3.1. Differential Constraint Equations, Aperture Problem.

All variational optical flow approaches impose an invariance assumption on some feature vector $g(x,t) \in \mathbb{R}^p$, derived from an image sequence $f(x,t)$ as discussed in Sect. 2.1. Under perfect conditions, any point moving along the trajectory $x(t)$ over time t with speed $u(x,t) := \frac{\mathrm{d}}{\mathrm{d}t}x(t)$ does not change its appearance, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t} g(x(t),t) = 0\,. \tag{3.2}$$

Without loss of generality, motion at $t = 0$ is considered only in what follows. Applying the chain rule and dropping the argument $t = 0$ for clarity, leads to the *linearized* invariance constraint,

$$\mathrm{J}_g(x)u(x) + \partial_t g(x) = 0\,. \tag{3.3}$$

Validity of this approximation is limited to displacements of about 1 pixel for real data as elaborated in Sect. 2.2.2, which seriously limits its applicability. However, Sect. 3.3.3 describes an approach to alleviating this restriction and thus for now it is safe to assume that the assumption is fulfilled.

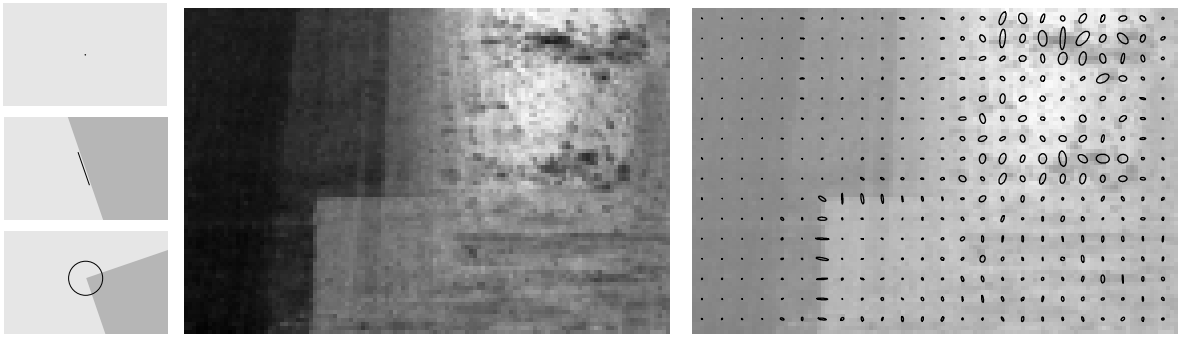
A least squares solution to (3.3) is given by $(S(x))^{-1}(\mathrm{J}_g^\top(x)(\partial_t g(x)))$ where

$$S(x) := \mathrm{J}_g^\top(x)\,\mathrm{J}_g(x). \tag{3.4}$$

However, in order to understand the actual information content of equation system (3.3), the locally varying properties of the Jacobian matrix $\mathrm{J}_g(x)$ have to be examined:

$\mathrm{rank}(\mathrm{J}_g) = 0$:	*void* constraints on $u(x)$ (for $g(x,0) = \text{const.}$);
$\mathrm{rank}(\mathrm{J}_g) = 1$:	*ill-conditioned* constraints, a single component of $u(x)$ is determined only;
$p = \mathrm{rank}(\mathrm{J}_g) = 2$:	*unique solution* $u(x) = -\,\mathrm{J}_g^{-1}(x)(\partial_t g(x))$;
$p > \mathrm{rank}(\mathrm{J}_g) = 2$:	*over-determined* and possibly conflicting constraints on $u(x)$, cf. Fig. 8.

In the case of gray-valued features $g(x) = f(x) \in \mathbb{R}$, (3.3) is referred to as the *linearized brightness*



(a) synthetic scenarios (b) real image data (c) local information content

FIGURE 8. Ellipse representation of $S = \mathrm{J}_g^\top \mathrm{J}_g$ as in (3.2) for a patch feature vector with $p \gg 2$ (see Sect. 3.3.2). **(a)** Three synthetic examples with J_g having (top to bottom) rank 0, 1 and 2, respectively. **(b)** Real image data with homogeneous (left) and textured (right) region, image edges and corner (middle). **(c)** Locally varying information content (see Sect. 3.1) of the path features extracted from (b).

constancy constraint and imposes *only one* scalar constraint on $u(x) \in \mathbb{R}^2$, in the direction of the image gradient $\mathrm{J}_g(x) = (\nabla g(x))^\top \neq 0$, i.e.

$$\left\langle \frac{\nabla g(x)}{\|\nabla g(x)\|}, u(x) \right\rangle = -\frac{\partial_t g(x)}{\|\nabla g(x)\|} \,. \tag{3.5}$$

This limitation which only allows to determine the *normal flow* component is referred to as the *aperture problem* in the literature.

Furthermore, for real data, invariance assumptions do not hold exactly and compliance is measured by the data term as discussed in Sect. 3.3. Section 3.4 addresses regularization terms which further incorporate regularity priors on the flow so as to correct for data inaccuracies and local ambiguities not resolved by (3.3).

3.2. **The Approach of Horn & Schunck.** The approach by Horn & Schunck [68] is described in the following due to its importance in the literature, its simple formulation and the availability of well understood numerical methods for efficiently computing a solution.

3.2.1. *Model.* Here the original approach [68], expressed using the variational formulation (3.1), is slightly generalized from gray-valued features $g(x) = f(x) \in \mathbb{R}$ to arbitrary feature vectors $g(x) \in \mathbb{R}^p$. Deviations from the constancy assumption in (3.3) are measured using a quadratic function $\rho_{\mathrm{D}} = \rho_2^2$, leading to

$$E_{\mathrm{D}}(u) = \frac{1}{2} \int_\Omega \rho_{\mathrm{D}} \left(\|\mathrm{J}_g(x) u(x) + \partial_t g(x)\|_F \right) \mathrm{d}x \,. \tag{3.6}$$

As for regularization, the quadratic length of the flow gradients is penalized using $\rho_{\mathrm{R}} = \rho_2^2$, to enforce smoothness of the vector field and to overcome ambiguities of the data term (e.g. aperture problem; see Sect. 3.1):

$$E_{\mathrm{R}}(u) = \frac{1}{2\sigma^2} \int_\Omega \rho_{\mathrm{R}}(\|\mathrm{J}_u(x)\|_F) \mathrm{d}x \,. \tag{3.7}$$

The only parameter $\sigma > 0$ weights the influence of regularization against the data term.

3.2.2. *Discretization.* Finding a minimum of $E(u) = E_{\mathrm{D}}(u) + E_{\mathrm{R}}(u)$ using numerical methods requires discretization of variables and data in time and space. To this end, let $\{x^i\}_{i\in[n]}$ define a regular two-dimensional grid in Ω, and let $g^1(x^i)$ and $g^2(x^i)$ be the discretized versions of $g(x,0)$ and $g(x,1)$ of the input image sequence, respectively. Motion variables $u(x^i)$ are defined on the same grid and stacked into a vector u:

$$u(x^i) = \begin{pmatrix} u_1(x^i) \\ u_2(x^i) \end{pmatrix}, \qquad u = \begin{pmatrix} (u_1(x^i))_{i\in[n]} \\ (u_2(x^i))_{i\in[n]} \end{pmatrix} \in \mathbb{R}^{2n}. \tag{3.8}$$

The appropriate filter for the discretization of the spatial image gradients $\partial_i g$ strongly depends on the signal and noise properties as discussed in Sect. 2.2.3. A recent comparison [115] reports that a 5-point derivative filter ($\frac{1}{12}\{-1,8,0,-8,1\}$) applied to $\frac{1}{2}(g^1+g^2)$ performs best. Temporal gradients are approximated as $\partial_t g(x^i) \approx g^2(x^i) - g^1(x^i)$.

As a result, the discretized objective function can be rewritten as

$$E(u) = \frac{1}{2}\|Du + c\|^2 + \frac{1}{2\sigma^2}\|Lu\|^2, \tag{3.9}$$

using the linear operators

$$D := \begin{pmatrix} D_{1,1} & D_{1,2} \\ \vdots & \vdots \\ D_{p,1} & D_{p,2} \end{pmatrix}, \quad c := \begin{pmatrix} c_1 \\ \vdots \\ c_p \end{pmatrix}, \quad L := \begin{pmatrix} L_{1,1} \\ L_{1,2} \\ L_{2,1} \\ L_{2,2} \end{pmatrix}, \tag{3.10}$$

with *data* derivatives $c_j := (\partial_t g_j(x^i))_{i\in[n]}$ and $D_{j,k} := \operatorname{diag}\left((\partial_k g_j(x^i))_{i\in[n]}\right)$. The matrix operator $L_{l,k}$ applied to *variable* u approximates the spatial derivative ∂_k of the flow component u_l using the 2-tap linear filter $\{-1,+1\}$ and Neumann boundary conditions.

3.2.3. *Solving.* Objective function (3.9) is strictly convex in u under mild conditions [109] and thus a *global* minimum of this problem can be determined by finding a solution to $\nabla_u E(u) = 0$. This condition explicitly reads

$$(D^\top D + \sigma^{-2} L^\top L)u = -D^\top c \tag{3.11}$$

which is a linear equation system of size $2n$ in $u \in \mathbb{R}^{2n}$ with a positive definite and sparse matrix. A number of well-understood iterative methods exist to efficiently solve this class of problems even for large n [104].

3.2.4. *Examples.* Figure 9 illustrates the method by Horn & Schunck for a small synthetic example. The choice of parameter σ is a trade-off between smoothing out motion boundaries (see Fig. 9(b)) in the true flow field (Fig. 9(a)) and sensitivity to noise (Fig. 9(d)).

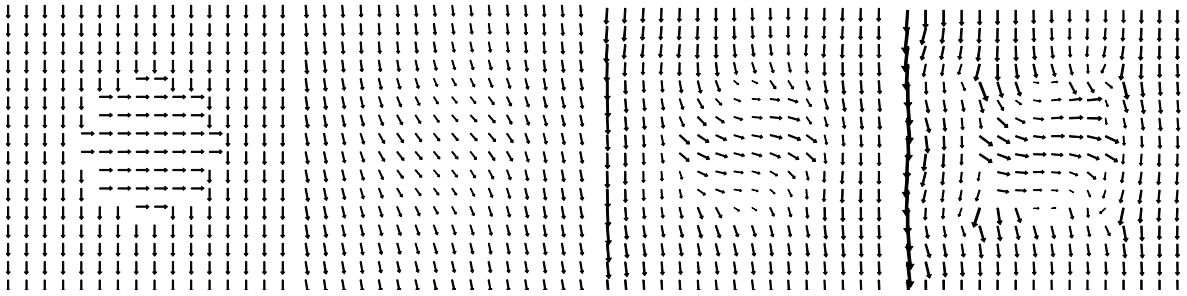
3.2.5. *Probabilistic Interpretation.* Considering $E(u)$ as a the log-likelihood function of a probability density function gives rise to the maximum a-posteriori interpretation of the optimization problem (3.1), i.e.

$$\sup_{u\in U} \mathrm{p}(u\,|g,\sigma)\,, \quad \mathrm{p}(u\,|g,\sigma) \propto \exp(-E(u))\,. \tag{3.12}$$

As $E(u)$ is quadratic and positive definite due to the assumptions made in Sect. 3.2.3, this posterior is a Gaussian multivariate distribution

$$\mathrm{p}(u\,|g,\sigma) = \mathcal{N}(u;\mu,\Sigma) \tag{3.13}$$

with precision (inverse covariance) matrix $\Sigma^{-1} = D^\top D + \sigma^{-2} L^\top L$ and mean vector $\mu = -\Sigma^{-1} D^\top c$ that solves (3.11).



(a) correct flow (b) estimated flow, $\sigma^2 = 10^2$ (c) estimated flow, $\sigma^2 = 10^3$ (d) estimated flow, $\sigma^2 = 10^4$

FIGURE 9. **(a)** Synthetic flow field used to deform an image. **(b)–(d)** Flow field estimated by the approach by Horn & Schunck with decreasing strength of the smoothness prior.

Examining the conditional distribution of $u^i \in \mathbb{R}^2$ allows to quantify the sensitivity of u. To this end a permutation matrix $Q = \begin{pmatrix} Q_i \\ Q_{\bar{i}} \end{pmatrix} \in \mathbb{R}^{2n \times 2n}$, $Q^\top Q = I$, is defined such that $u^i = Q_i u$. Then, fixing $Q_{\bar{i}} u = Q_{\bar{i}} \mu$ leads to

$$\mathrm{p}\big(u^i \,\big|\, Q_{\bar{i}} u\big) = \mathcal{N}\Big(\hat{\mu}^i, \hat{\Sigma}^i\Big) \tag{3.14}$$

with $\hat{\mu}^i = Q_i \mu$ and

$$\hat{\Sigma}^i = Q_i \Sigma Q_i^\top - (Q_{\bar{i}} \Sigma Q_i^\top)(Q_{\bar{i}} \Sigma Q_{\bar{i}}^\top)^{-1}(Q_i \Sigma Q_{\bar{i}}^\top)\,. \tag{3.15}$$

Using the matrix inversion theorem to invert Σ block-wise according to Q and restricting the result to u_i, reveals

$$Q_i \Sigma^{-1} Q_i = \Big(Q_i \Sigma Q_i^\top - (Q_{\bar{i}} \Sigma Q_i^\top)(Q_{\bar{i}} \Sigma Q_{\bar{i}}^\top)^{-1}(Q_i \Sigma Q_{\bar{i}}^\top)\Big)^{-1}\,. \tag{3.16}$$

Comparison of (3.15) to (3.16) and further analysis yields (for non-boundary pixels)

$$\hat{\Sigma}^i = \big(Q_i \Sigma^{-1} Q_i\big)^{-1} = \big(S^i + 4\sigma^{-2} I\big)^{-1} \tag{3.17}$$

with $S^i = S(x^i)$ as defined by (3.4). Consequently, smaller values of σ reduce the sensitivity of u^i, but some choice $\sigma > 0$ is inevitable for singular S^i.

3.3. **Data Terms.**

3.3.1. *Handling Violation of the Constancy Assumption.* The data term as proposed by Horn & Schunck was refined and extended in literature in several ways with the aim to cope with the challenging properties of image data of real applications, see Sect. 2.2.5.

Changes of the camera viewpoint as well as moving or transforming objects may cause previously visible image features to disappear due to occlusion, or vice versa to emerge (dis-occlusion), leading to discontinuous changes of the *observed* image features $g(x(t), t)$ over time and thus to a violation of the invariance constraint (3.2).

Surface reflection properties like specular reflections that vary as the viewpoint changes, and varying emission or illumination (including shadows) also cause appearance to change, in particular in natural and outdoor scenes.

With some exceptions, most approaches do not to explicitly model these cases and instead replace the quadratic distance function ρ_2^2 by the convex ℓ^2-distance or its differentiable approximation $\rho_{2,\epsilon}$, to reduce the impact of outliers in regions with strong deviation from the invariance assumption. A

number of *non-convex* alternatives have been proposed in the literature, including the truncated square distance $\rho_{2,\lambda}$, which further extend this concept and are often referred to as "robust" approaches.

Another common method is to replace the constancy assumption on the image brightness by one of the more complex feature mappings $g(x,t)$ introduced in Sect. 2.1, or combinations of them. The aim is to gain more descriptive features that overcome the ambiguities described in Sect. 3.1, e.g. by including color or image structure information from a local neighborhood. Furthermore, robustness of the data term can be increased by choosing features invariant to specific image transformations. For example, $g(x) = \nabla f(x)$ is immune to additive illumination changes.

3.3.2. *Patch Features.* Contrary to the strongly localized brightness feature $g(x) = f(x)$, local image patches sampled from a neighborhood $\mathcal{N}(x)$ of x,

$$g(x^i,t) = \big(f(x^j,t)\big)_{x^j\in\mathcal{N}(x^i)} \in \mathbb{R}^p, \quad p = |\mathcal{N}(x^i)| \tag{3.18}$$

provide much more reliable information on u in textured image regions. In fact, *local* approaches set $E_{\mathrm{R}}(u) = 0$ and rely only the information contained in the data term.

The most prominent instance introduced by Lucas & Kanade [84], chooses a Gaussian weighted quadratic distance function,

$$\rho^2_{w^i}(z) := \|\operatorname{diag}(w^i)^{\frac{1}{2}} z\|^2, \quad w^i := (w(x^i - x^j))_{x^j\in\mathcal{N}(x^i)} \tag{3.19}$$

and $w(x) := \exp\left(-\|x\|^2/(2\sigma^2)\right)$. Solving the variational problem (3.1) decomposes into n linear systems of dimension 2 each. Furthermore, the sensitivity in terms of (3.17) reduces to $\hat{\Sigma}^i = (S^i)^{-1}$ and

$$S^i = \sum_{x^j\in\mathcal{N}(x^i)} w(x^i - x^j)\left(\mathrm{J}_g^\top(x^j)\,\mathrm{J}_g(x^j)\right) \tag{3.20}$$

equals the so-called *structure tensor*. At locations with numerically ill-conditioned J_g, cf. Fig. 8 and the discussion in Sect. 3.1, no flow can be determined reliably which leads to possibly sparse results. The works [111, 30] overcome this drawback by complementing this data term by a regularization term.

3.3.3. *Multiscale.* As discussed in Sect. 2.2.3, the range of displacements $u(x)$ that can be accurately estimated, is limited to about 1 pixel which does not conform to the larger magnitude of motion fields typically encountered in practical applications. Multiscale methods allow to remove this restriction to some extent. They implement a *coarse-to-fine* strategy for approximately determining large displacements on spatially band-limited image data and complementing flow details on finer scales.

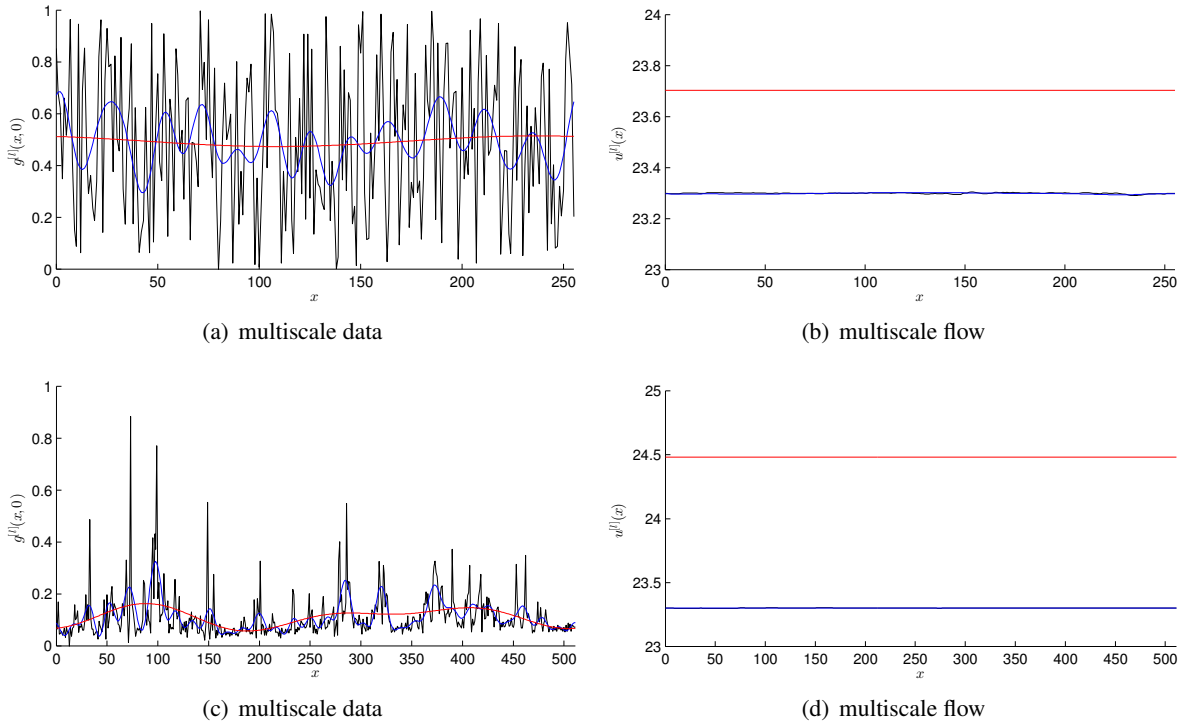
The underlying idea is introduced by means of a multiscale representation $\{g^{[l]}\}_{l\in[n_l]}$ of image data, where $l = 0$ and $l = n_l - 1$ refer to the finest and coarsest scale, respectively. More precisely, $g^{[l]}$ is a spatially band-limited version of g with $\omega_{x,\max} < s_l\pi$ with $1 = s_0 > s_1 \cdots s_{n_l-1} > 0$. The computation is described by the following recursive scheme with $u^{[n_l]}(x) = 0$:

- $g^{[l]}(x,t) := h_l * g(x + t\cdot u^{[l+1]}, t)$
- $\delta u^{[l]} := \arg\min_u E(u)$ on data $g^{[l]}(x,t)$
- $u^{[l]}(x) := u^{[l+1]}(x) + \delta u^{[l]}(x)$

with a suitable approximation of the ideal low-pass filter h_l with frequency response

$$\hat{h}_l(\omega_x,\omega_t) \approx \begin{cases} 1 & \|\omega_x\|_\infty < s_l\pi \\ 0 & \text{otherwise} \end{cases}. \tag{3.21}$$

Figure 10 demonstrates the method for two simple examples.



(a) multiscale data (b) multiscale flow

(c) multiscale data (d) multiscale flow

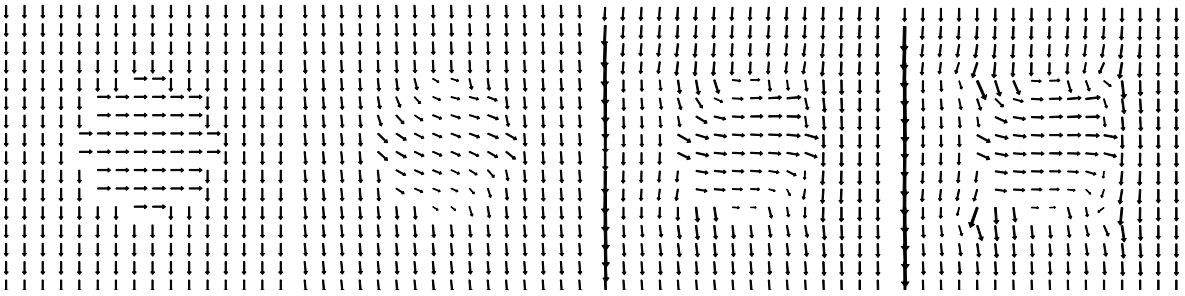
FIGURE 10. Multiscale flow estimation: **(a)** An image (white noise) $g^{[l]}(x,0)$ represented at multiscale levels $l = 0$ (black), $l = 3$ (blue), $l = 6$ (red) with $s_l = 2^{-l}$, i.e. band-limited to $s_l\pi$. **(b)** Estimate $u^{[l]}$ (same color encoding as in (a)) of correct constant flow $u(x) = 23.3$ on multiscale level l. **(c)-(d)** Same as (a)-(b) for a single line of a real image (©LaVision GmbH) as found in particle image velocimetry, an optical fluid flow estimation method.

Actual implementations further make use of the band-limited spectrum of the filtered data and subsample the data according to the Nyquist-Shannon sampling theorem, leading to a data representation referred to as *resolution pyramid*. The recursive structure allows in turn to approximate h_l by chaining filters with small support for computational efficiency.

3.4. **Regularization.** Ill-posed data terms, sensor noise and other distortions lead to *sparse* and *locally inaccurate* flow estimates. Variational approaches allow to incorporate priors on the motion regularity by means of additional terms $E_{\mathrm{R}}(u)$. For suitable models $E_{\mathrm{D}}(u)$ and $E_{\mathrm{R}}(u)$, accuracy profits from this concept as the global solution to minimization problem (3.1) represents the best flow field according to both observations and priors. Furthermore, in contrast to local methods, *missing* flow information is approximately inferred according to the smoothness prior. This is in particular essential in connection with ill-posed data terms (cf. Sect. 3.1).

3.4.1. *Regularity Priors.* A number of a-priori constraints $u \in U$ for flow estimation have been proposed in the literature, based on prior knowledge specific to the application domain. Examples include

- inherent geometrical constraints induced by multi-camera setups (Sect. 3.5.2),
- physical properties of flows in experimental fluid mechanics (Sect. 3.5.3).



(a) correct flow (b) estimated flow, $\sigma^2 = 10^2$ (c) estimated flow, $\sigma^2 = 10^3$ (d) estimated flow, $\sigma^2 = 10^4$

FIGURE 11. **(a)** Synthetic flow field used to deform an image. **(b)–(d)** Flow field estimated by the approach by Horn & Schunck, however with $\ell_1 - TV$-regularization, with decreasing strength of the smoothness prior.

Formally, strict compliance with a constraint $u \in U$ can be incorporated into the variational formulation (3.1) by means of the corresponding indicator function

$$E_{\mathrm{R}}(u) = \delta_U(u)\,. \tag{3.22}$$

In many applications, however, the set U cannot be specified precisely. Then a common approach is to replace δ_U by a smoother function measuring the distance of u to U in some sense,

$$E_{\mathrm{R}}(u) = \rho(u - \Pi_U u). \tag{3.23}$$

For example, the regularization term of the Horn & Schunck approach presented in Sect. 3.2 may be written as

$$E_{\mathrm{R}}(u) = \|Lu\|^2 = \|u - \Pi_{\ker(L)}(u)\|_L^2 \tag{3.24}$$

with semi-norm $\|x\|_L := \|Lx\|$ and set $U = \ker(L)$. Generalizations of the approach of Horn & Schunck are based on the same L and modify the distance function (Sect. 3.4.2) or refine them to become locally adaptive and anisotropic (Sect. 3.4.3).

Further extensions replace the gradient operator in (3.7) and its discretization L by other operators having a larger space $U = \ker(L)$. For example, operators involving second order derivatives $\nabla\,\mathrm{div}$ and $\nabla\,\mathrm{curl}$ have been used for flow estimation in experimental fluid dynamics [141, 142, 143] (cf. Sect. 3.5.3).

3.4.2. *Distance Functions.* Occlusion of objects do not only lead to sudden changes of the projected appearance (cf. Sect. 3.3), but also to *motion discontinuities* whose preservation during flow estimation is crucial in many applications and for the interpretation of videos. The penalization of large motion gradients J_u can be reduced by replacing the quadratic distance function ρ_2^2 in (3.7) by convex or non-convex alternatives, see (2.3) for some examples.

Figure 11 demonstrates the effect of replacing the quadratic distance measure of the approach by Horn & Schunck (Sect. 3.2) by $\rho_{\mathrm{R}} = \rho_2$. It becomes apparent that motion discontinuities can be better resolved than with $\rho_{\mathrm{R}} = \rho_2^2$ (see Fig. 9).

3.4.3. *Adaptive, Anisotropic and Non-local Regularization.* A further option is to include a-priori information on the *location* and *alignment* of motion discontinuities by using a spatially varying, *adaptive* and possibly *anisotropic* norm in (3.7),

$$E_{\mathrm{R}}(u) = \sigma^{-2} \int_\Omega \rho_{\mathrm{R}}(\|\,\mathrm{J}_u(x)\|_{W(x)})\mathrm{d}x\,, \tag{3.25}$$

with $\|A\|_W := \|AW\|_F$ and (omitting the dependency on x)

$$W = \begin{pmatrix} w_1 e^1 & w_2 e^2 \end{pmatrix} . \tag{3.26}$$

The normalized orthogonal directions $e^1, e^2 \in \mathbb{R}^2$ point *across* and *along* the assumed motion boundary, respectively. The positive eigenvalues w_1 and w_2 control relative penalization of flow changes in the according direction.

A common assumption made in literature, e.g. [134, 129], is that image edges and flow discontinuities coincide and to facilitate changes of $u(x)$ *across* the assumed boundary e_1. For general features $g(x)$, the notion of image edge is here defined by choosing e^1 and e^2 as the normalized direction e of the largest and smallest change of $\|\,\mathrm{J}_g\, e\|$, respectively, given by the eigenvectors of $S = \mathrm{J}_g^\top\, \mathrm{J}_g$. The associated eigenvalues $\lambda_1 \geq \lambda_2 \geq 0$ of S control the strength of smoothness by setting $w_i = 1 - \rho(\lambda_i)$, $i = 1, 2$ and suitable increasing $\rho(x) \in [0, 1]$ with $\rho(0) = 0$. This defines an *anisotropic* and *image-driven* regularization. Note that for the gray-valued case $g(x) = f(x) \in \mathbb{R}$ the formulation simplifies to $e^1 = \|\nabla g\|^{-1} \nabla g$, $\lambda_1 = \|\nabla g\|^2$ and $\lambda_2 = 0$. The class of *flow-driven* approaches replace the dependency on $g(x)$ of the terms above by the flow $u(x)$ to be estimated. This nonlinear dependency can be taken into account without compromising convexity of the overall variational approach [129].

While the approaches so far measure *locally* the regularity of flows u, approaches such as [79] adopt non-local functionals for regularization developed in other contexts [76, 48, 44] for optical flow estimation. Regularization is then more generally based on the similarity between *all* pairs $(u(x), u(x'))$ with $x, x' \in \Omega$, weighted by mutual position and feature distances.

3.5. **Further Extensions.** Three extensions of the basic variational approach are sketched: a natural extension of spatial regularizers to the spatio-temporal domain (Sect. 3.5.1), regularization based on the two-view geometry (cf. Sect. 2.3) and relative rigid motions for computer vision applications (Sect. 3.5.2) and a case study of PDE-constrained variational optical flow estimation in connection with imaging problems in experimental fluid dynamics (Sect. 3.5.3).

3.5.1. *Spatio-Temporal Approach.* The preceding discussion reduced the motion estimation problem to determining displacements between *two* image frames only and thus ignored consistencies of the flow *over time*. Although in many applications recording rates are fast compared to dynamical changes due to modern sensors, only few approaches exploit this fact by introducing *temporal smoothness priors*.

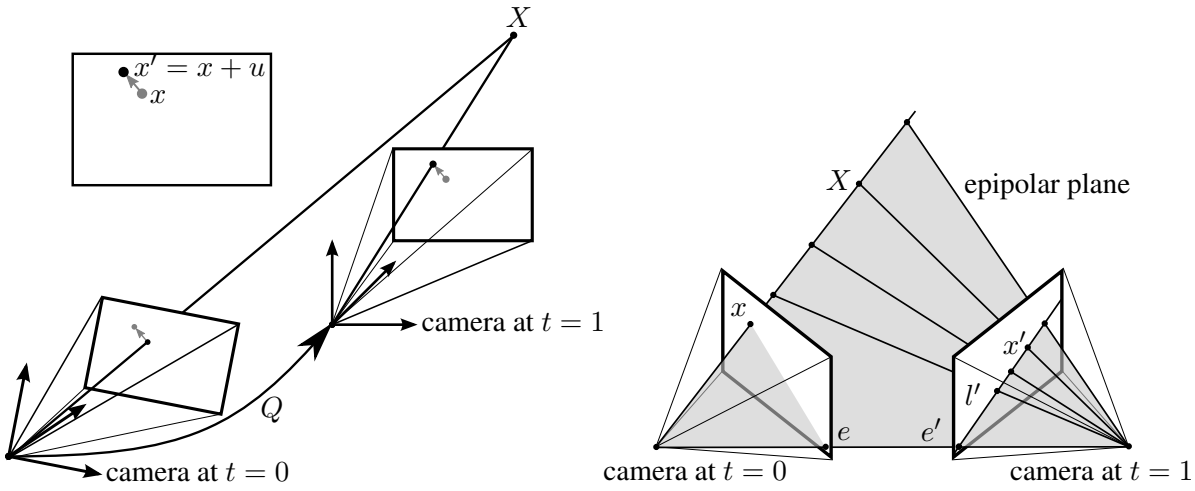
The work [130] proposed to process a batch of image frames *simultaneously* and to extend the flow field domain along the time axis $u : \Omega \times [0, T] \mapsto \mathbb{R}^2$. While data terms are independently imposed for each time t, the smoothness prior is extended by a temporal component to

$$E_{\mathrm{R}}(u) := \int_{\Omega \times [0,T]} \rho_{\mathrm{R}}(\|\,\mathrm{J}_{u,t}(x,t)\|_W) \mathrm{d}x \mathrm{d}t\,. \tag{3.27}$$

Here, $\mathrm{J}_{u,t}$ represents the spatio-temporal derivatives and $\rho_{\mathrm{R}} \|\cdot\|_W$ is a three-dimensional extension of the *anisotropic, flow-driven* distance function discussed in Sect. 3.4.3. It allows to account for *small* position changes of moving objects between consecutive frames within the support of the regularization term (≤ 1 px) by supporting smoothness *along* an assumed trajectory.

Larger displacements, however, require matching of temporally associated regions e.g. using a multiscale framework (Sect. 3.3.3) but then enable to regularize smoothness of *trajectories* over multiple frames as proposed in [126].

Online methods are an appealing alternative whenever processing a batch of image frames is not feasible due to resource limitations. This approach is addressed in Sect. 5.3.



(a) optical flow induced by camera motion (b) epipolar constraint

FIGURE 12. **(a)** Relative motion $Q \in \mathsf{SE}(3)$ of the camera w.r.t. a world coordinate system causes the projection x of a static scene point $X = z(x)y$ to move from x to $x' = x + u(x)$ in the image plane. **(b)** Any two projections x, x' of a scene point X are related by the essential matrix E as in (3.29), defining an *epipolar plane* and their projection, the *epipolar lines* defined by $\{x, e\}$ and $\{x', e'\}$ in the image plane at $t = 0$ and $t = 1$, respectively.

3.5.2. *Geometrical Prior Knowledge.* In applications with a perspective camera as image sensor, the geometrical scene structure strongly determines the observed optical flow (Sect. 2.3). This section briefly addresses the most common assumptions made and the constraints that follow.

Often, a *static scene* assumption is made, meaning that all visible scene points have zero velocity with respect to a world coordinate system. Then the observed motion is only induced by the camera moving in the scene. Using the notation introduced in Sect. 2.3, the camera motion is denoted by $Q \in \mathsf{SE}(3)$ (cf. Fig. 12(a)), parametrized by rotation $R \in \mathsf{SO}(3)$ and translation $h \in \mathbb{R}^3$, so that any scene point $Y \in$ is transported to $Y' \simeq Q^{-1}Y$.

The following discussion of common setups and their implications on the observed motion implicitly assumes that the scene point is visible in both frames. Using assumption (2.41) for the internal camera parameters allows to work with normalized coordinates (2.32). The point corresponding to x is denoted by x', due to (2.42).

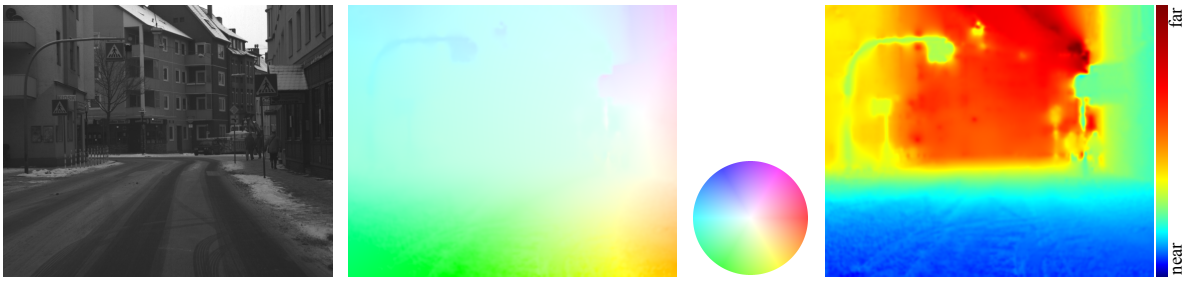
Static scene, general motion: Let the depth map $z(x) : \Omega \mapsto \mathbb{R}$ parametrize the scene point $X := z(x) \left(\begin{smallmatrix} x \\ 1 \end{smallmatrix}\right)$ visible at x in the camera plane in the first frame. Then the projected correspondences are given in homogeneous coordinates by

$$y' \simeq PQ^{-1}Y = R^\top \left(z(x) \left(\begin{smallmatrix} x \\ 1 \end{smallmatrix}\right) - h \right) , \tag{3.28}$$

see Fig. 12(a) for an illustration. Figure 13 shows the optical flow field $u(x)$ conforming to constraint (3.28) for a real application.

It is possible to eliminate the dependency on $z(x)$, that typically is unknown, by means of the *essential matrix* $E := R^\top [h]_\times$, leading to the *epipolar constraint*

$$(y')^\top E y = 0 , \tag{3.29}$$



(a) frame from a real image sequence (b) estimated flow $u(x)$ and color code (c) estimated depth $z(x)$

FIGURE 13. **(a)** A single frame from an image sequence recorded by a camera moving forward through an approximately static scene. **(b)** Optical flow estimated using the parametrization $u(x) = u(x; Q, z(x))$ according to (3.28) and global optimization for $Q \in \mathsf{SE}(3)$, $z \in \mathbb{R}^n$, see [18] for details. Displacement length and direction are encoded by saturation and hue, respectively, see color code on the right. **(c)** Estimated depth parameter $z(x)$ using the color code on the right. Scene structure is more evident in this representation and therefore the spatial smoothness prior on the flow was formulated as regularization term on the *depth* $z(x)$ instead of displacements $u(x)$.

as illustrated by Fig. 12(b). This gives rise to an orthogonal decomposition [112] of an observed correspondence $\hat{x}'$ into

$$\hat{x}' = \hat{x}'_e + \hat{x}'_\perp \tag{3.30}$$

with $\hat{x}'_e$ fulfilling (3.29) and orthogonal deviations $\hat{x}'_\perp$.

Even without knowing a-priori (R, h), equation (3.29) provides a valuable prior: Valgaerts et al. [121] propose joint computation of the fundamental matrix F related to E by (2.40) and optical flow constrained via (3.29). They show that estimation of F is more stable and that flow accuracy is significantly increased.

Static scene, coplanar camera motion: If the camera translates parallel to the image plane only, i.e. $R = I$ and $h = \left(\begin{smallmatrix} b \\ 0 \end{smallmatrix}\right)$ with $b \in \mathbb{R}^2$, the observed flow is constrained to a locally varying *one-dimensional* subspace parametrized by the inverse depth,

$$u(x) = z^{-1}(x) b\,. \tag{3.31}$$

Stereoscopic camera setups fulfill the static scene assumption as they can be interpreted as an instantaneous camera motion with baseline $\|b\|$. For details see e.g. [27].

Planar and static scene, general camera motion: In applications where the scene can be (locally) approximated by a plane such that $\langle n, X \rangle - d = 0$ for all space points X with plane parameters d, n as in (2.46), all correspondences fulfill

$$y' \simeq Hy,\; H = R^\top \left(I - \frac{1}{d} h n^\top \right)\,, \tag{3.32}$$

where $H \in \mathbb{R}^{3\times 3}$ defines a *homography* – cf. Eq. (2.49).

3.5.3. *Physical Prior Knowledge.* Imaging of dynamic phenomena in natural sciences encounters often scenarios where physical prior knowledge applies. Examples include particle image velocimetry [3] or Schlieren velocimetry [8], where the motion of fluids is observed that is governed by physical

laws. While local methods such as cross-correlation methods are commonly used to evaluate the obtained image sequences [135, 3], variational approaches [100, 101, 61] provide a more appropriate mathematical framework for exploiting such prior knowledge and the estimation of physically consistent optical flows.

For instance the Helmholtz decomposition of vector fields enables to define regularizers in terms of higher-order partial flow derivatives in a natural way [141, 142, 143]. Constraints like incompressibility can be enforced as hard or soft contraints using advanced methods of convex programming, to cope with imaging imperfections. Conversely, flow field estimates obtained by other image processing methods can be denoised so as to restore physically relevant structure [125].

A particularly appealing approach exploits directly some equation from fluid dynamics, that governs the flow as state of the physical system which is observed through an imaging sensor [102, 103]. The state is regarded as hidden and only observable through the data of an image sequence that depicts the velocity of some tracer suspended in the fluid. The variational approach of fitting the time varying state to given image sequence data results in a PDE-constrained optimization or distributed parameter control problem, respectively.

As example the approach [102] is sketched based on the Stokes system

$$-\mu\Delta u + \nabla p = f_\Omega \qquad \text{in } \Omega, \tag{3.33a}$$

$$\operatorname{div} u = 0 \qquad \text{in } \Omega, \tag{3.33b}$$

$$u = f_{\partial\Omega} \qquad \text{on } \partial\Omega, \tag{3.33c}$$

that for given $f_\Omega, f_{\partial\Omega}$ with $\int_{\partial\Omega}\langle n, f_{\partial\Omega}\rangle \mathrm{d}s = 0$ (n denotes the outer unit normal of the Lipschitz domain Ω) has a unique solution u, p under classical assumptions [49, Ch. I]. Here $f_\Omega, f_{\partial\Omega}$ are not regarded as given data but as control variables, to be determined so that the flow u not only satisfies (3.33) but fits also given image sequence data. To achieve the latter, both the state variables u, p and the control variables $f_\Omega, f_{\partial\Omega}$ are determined by minimizing in the two-dimensional case $d = 2$ the objective

$$E(u, p, f_\Omega, f_{\partial\Omega}) = E_{\mathrm{D}}(u) + \alpha \int_\Omega \rho_2^2(f_\Omega)\mathrm{d}x + \gamma \int_{\partial\Omega} \rho_2^2\big(\langle n^\perp, \nabla f_{\partial\Omega}\rangle\big)\mathrm{d}s, \qquad \alpha, \gamma > 0. \tag{3.34}$$

The first term $E_{\mathrm{D}}(u)$ denotes a data term of the form (3.6), and the remaining two terms regularize the control variables so as to make the problem well-posed.

For related mathematical issues (e.g. constraint qualification and existence of Lagrange multipliers) see [53, Ch. 6] and [71, Ch. 1], and furthermore [54, 53] for related work outside the field of mathematical imaging based on the general Navier-Stokes system.

3.6. **Algorithms.** The choice of an optimization method for numerically minimizing the functional (3.1) depends on the specific formulation of the terms E_{D} and E_{R} involved. Suitable methods can be broadly classified into

- algorithms for minimizing *smooth convex* functionals,
- algorithms for minimizing *non-smooth convex* functionals,
- algorithms for locally minimizing *non-convex* functionals.

In view of the typical multiscale implementation of the data term (Sect. 3.3.3) that enables a quadratic approximation at each resolution level, this classification is applied to the regularizer E_{R} only and each class is discussed in turn in the sections to follow. The reader should note that convex non-quadratic data terms, as discussed in Sect. 3.3.1, can be handled in a similar way as the convex non-smooth regularizer below, and a number of closely related alternatives exist (e.g. [36]). Since convex programming has been extensively studied in the literature, the following presentation is confined to

representative case studies that illustrate in each case the underlying idea and application of a general principle.

3.6.1. *Smooth Convex Functionals.* It is useful to distinguish quadratic and non-quadratic functionals. The approach of Horn and Schunck (Sect. 3.2) is a basic representative of the former class. Solving the corresponding linear positive definite sparse system can be efficiently done by established methods [104]. More sophisticated implementations are based on numerical multigrid methods [26]. These are optimal in the sense that runtime complexity $O(n)$ linearly depends on the problem size n. Dedicated implementations run nearly at video frame rate on current PCs.

For more general data-dependent quadratic regularizers and especially so for non-quadratic convex regularizers (cf. Sect. 3.4.3 and [129]), multigrid implementation that achieve such runtimes require some care. See [28, 29, 55] for details and to [119] for a general exposition.

3.6.2. *Non-smooth convex functionals.* This class of optimization problems has received considerable attention in connection with mathematical imaging, inverse problems, machine learning and in other fields during the recent years, due to the importance of non-smooth convex sparsity enforcing regularization. See [12] for a recent overview.

The *total variation* regularizer

$$\begin{aligned} E_{\mathrm{R}}(u) = \mathrm{TV}(u) &:= \sup_{v \in \mathcal{D}} - \int_{\Omega} \langle u, \operatorname{Div} v \rangle \mathrm{d}x, \\ \mathcal{D} &:= \{ v \in C_0^{\infty}(\Omega; \mathbb{R}^d)^d \colon \|v(x)\|_F \leq 1, \ \forall x \in \Omega \}, \\ \operatorname{Div} v &= (\operatorname{div} v^1, \ldots, \operatorname{div} v^d)^{\top} \end{aligned} \tag{3.35}$$

is a basic representative of the class of non-smooth convex functionals and appropriate to expose a general strategy of convex programming that is commonly applied: problem splitting into subproblems for which the proximal mapping can be efficiently evaluated.

The simplest *anisotropic* discretization of (3.35) that is particularly convenient from the viewpoint of convex programming, reads

$$\sum_{ij \in E(G)} \sum_{k \in [d]} |u_k(x^i) - u_k(x^j)|, \tag{3.36}$$

where $\{x^i\}_{i \in [n]}$ are the locations indexed vertices $V = [n]$ of a grid graph $G = (V, E)$ in Ω, and $E = E(G)$ are the corresponding edges connecting adjacent vertices resp. locations along the coordinate axes. Defining the vector

$$z \in \mathbb{R}^{d \times |E(G)|}, \quad z_{k,ij} = u_k(x^i) - u_k(x^j) \tag{3.37}$$

leads to the reformulation of (3.36)

$$\|z\|_1, \qquad Lu = z \tag{3.38}$$

where the linear system collects all equations of (3.37). As a consequence, the overall discretized problem reads

$$\min_{u,z} E_{\mathrm{D}}(u) + \alpha \|z\|_1 \quad \text{subject to} \quad Lu - z = 0, \qquad \alpha > 0 \tag{3.39}$$

to which the ADMM approach [25] can be applied, that entails a sequence of partial minimizations of the augmented Lagrangian corresponding to (3.39),

$$L_{\lambda}(u, z, w) = E_{\mathrm{D}}(u) + \alpha \|z\|_1 + \langle w, Lu - z \rangle + \frac{\lambda}{2} \|Lu - z\|^2. \tag{3.40}$$

Specifically, with some parameter value $\lambda > 0$ and multiplier vector w, the three-steps iteration

$$u^{k+1} = \operatorname*{argmin}_u E_{\mathrm{D}}(u) + \langle w^k, Lu\rangle + \frac{\lambda}{2}\|Lu - z^k\|^2, \tag{3.41a}$$

$$z^{k+1} = \operatorname*{argmin}_z \alpha\|z\|_1 - \langle w^k, z\rangle + \frac{\lambda}{2}\|Lu^{k+1} - z\|^2, \tag{3.41b}$$

$$w^{k+1} = w^k + \lambda(Lu^{k+1} - z^{k+1}), \tag{3.41c}$$

is iteratively applied for $k = 0, 1, 2, \ldots$, with arbitrary initializations z^0, q^0, until a suitable termination criterion is met [25, Section 3.3.1].

Assuming a quadratic form or approximation of $E_{\mathrm{D}}(u)$ at some resolution level (Sect. 3.3.3), subproblem (3.41a) amounts to solve a sparse positive definite linear system similar to the basic approach of Horn & Schunck, to which a multigrid solver can be applied as discussed above. Subproblem (3.41b) amounts to computing the proximal mapping for the ℓ^1-norm and hence to perform a simple shrinkage operation. See [37, 93] for corresponding surveys.

3.6.3. *Non-convex functionals.* Similar to the preceding non-smooth convex case, approaches are of interest that can be conducted by solving a sequence of *simple* subproblems efficiently. Clearly, convergence to a *local* minimum can be only expected. In contrast to the simpler convex cases above, the absence of parameters is preferable that would have to be set properly, to ensure convergence to some local minimum for any initialization. For example, Lipschitz constants of gradients are rarely known in practice, and setting corresponding parameters savely enough will unduly slow down convergence even for smooth problems.

A general strategy will be outlined next and its application to the non-convex extension of the regularizer (3.36), using the distance function (2.3f),

$$\sum_{ij\in E(G)} \rho_{2,\lambda}\big(u(x^i) - u(x^j)\big). \tag{3.42}$$

In order to illustrate graphically the non-convexity of this regularizer from the viewpoint of optimization, consider three summands of the "fully" anisotropic version of (3.42),

$$\sum_{ij\in E(G)} \sum_{k\in[d]} \rho_{2,\lambda}\big(u_k(x^i) - u_k(x^j)\big). \tag{3.43}$$

defined on edges that meet pairwise in a common vertex,

$$\rho_{2,\lambda}\big(u_k(x^{i_1}) - u_k(x^{i_2})\big) + \rho_{2,\lambda}\big(u_k(x^{i_2}) - u_k(x^{i_3})\big) + \rho_{2,\lambda}\big(u_k(x^{i_3}) - u_k(x^{i_4})\big). \tag{3.44}$$

Setting for simplicity and w.l.o.g. $u_k(x^{i_1}) = u_k(x^{i_4}) = 0$ to obtain a function of two variables $u_k(x^{i_2})$, $u_k(x^{i_3})$, results in the corresponding graph depicted by Fig. 14. It illustrates the presence of many non-strict local minima and that the design of a convergent minimization algorithm is not immediate.

Next consider a single summand $\rho_{2,\lambda}(z_i - z_j)$ of (3.43) with two scalar variables denoted by z_i and z_j for simplicity. This function can be decomposed into the difference of two proper, lower-semicontinuous (lsc), convex functions g and h,

$$\begin{aligned}\rho_{2,\lambda}(z_i - z_j) &= \tau(z_i - z_j)^2 - \Big(\tau(z_i - z_j)^2 - \rho_{2,\lambda}(z_i - z_j)\Big)\\ &=: g(z_i, z_j) - h(z_i, z_j), \qquad \tau > 1.\end{aligned} \tag{3.45}$$

Applying this decomposition to each term of (3.43) yields

$$g(u) - h(u) \tag{3.46}$$

with $g(u) = \tau\|Lu\|^2$ as in (3.9), and with $h(u)$ equal to the sum of all edge terms of the form $h(u_k(x^i), u_k(x^j))$, $ij \in E$, $k \in [d]$, given by (3.45).

DC-programming (DC stands for Difference-of-Convex functions [69]) amounts to locally minimize (3.46) by solving a sequence of convex problems, defined by the closed affine majorization of the concave part $-h$,

$$u^{k+1} = \operatorname*{argmin}_u g(u) - \big(h(u^k) + \langle v^k, u - u^k\rangle\big), \qquad v^k \in \partial h(u^k), \tag{3.47}$$

where $\partial h(u^k)$ denotes the subdifferential of h at u^k. This two-step iteration in terms of (u^k, v^k) converges under mild conditions [67]. Smoothing the problem slightly by replacing the distance function $\rho_{2,\lambda}$ by $\rho_{2,\lambda,\varepsilon}$ defined by (2.3g), and replacing accordingly h by h_ε, yields $v^k = \nabla h_\varepsilon(u^k)$ and hence turns (3.47) into the sequence of problems

$$u^{k+1} = \operatorname*{argmin}_u g(u) - \langle \nabla h_\varepsilon(u^k), u - u^k\rangle. \tag{3.48}$$

Taking additionally into account the data term $E_{\mathrm{D}}(u)$ and assuming it (or its approximation) has quadratic form at some resolution level (Sect. 3.3.3), solving (3.48) amounts to a sequence of Horn & Schunck type problems to which numerical multigrid can be applied, due to the simple form $g(u) = \tau\|Lu\|^2$. Not any single parameter, e.g. for selecting the stepsize, has to be set in order to ensure convergence, and available code for a variational method can be directly applied. The price to pay for this convenience is a moderate convergence rate.

Figure 14 illustrates the beneficial effect of smoothing and robustness of the non-convex regularizer: Only the components of points $\binom{z_2}{z_3}$ that are close enough to the data $z_1 = u_k(x^{i_1}) = z_4 = u_k(x^{i_4}) = 0$, as specified by λ, are fitted to these data. For distant points with $z_2 \approx z_3$, regularization enforces $z_2 = z_3$ or does not affect them at all if $|z_2 - z_3|$ is large.

Applying the scheme (3.48) to (3.42) instead of (3.43) is straightforward. This does not affect $g(u)$ but merely ∇h_ε in (3.48), due to replacing the scalar variables in (3.45) by the corresponding vectors.

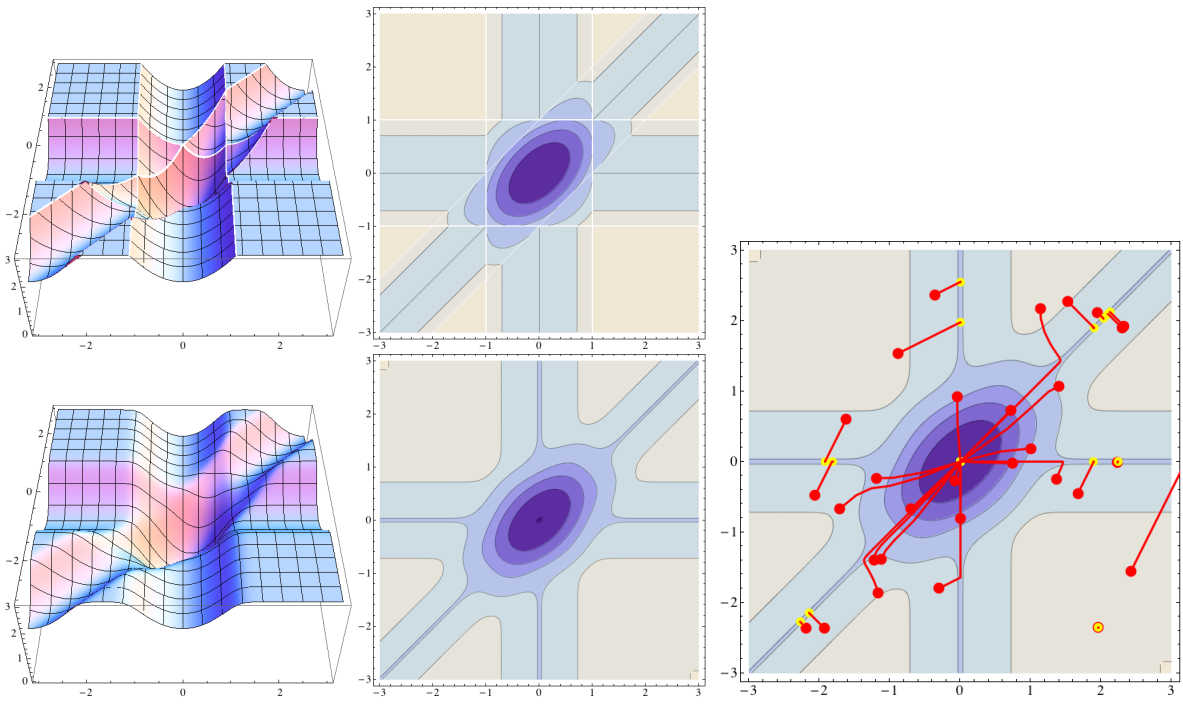


FIGURE 14. TOP ROW, LEFT: Two different illustrations of the non-convex, non-smooth objective (3.44). BOTTOM ROW, LEFT: The objective (3.44) smoothed by replacing the distance function $\rho_{2,\lambda}$ by $\rho_{2,\lambda,\varepsilon}$ with $\varepsilon = 0.2$, as defined by (2.3g). RIGHT PANEL: Sequences of iterates generated by (3.48) for 30 random points $(z_2, z_3)^\top$ (initial and final iterates are marked with red and yellow, respectively). The regularizer enforces fitting of the components z_2, z_3 to the data $z_1 = z_4 = 0$ as well as $z_2 = z_3$. It is robust in the sense that components that are too distant to either of these criteria, are not affected accordingly.

4. The Assignment Approach to Optical Flow Estimation

In this section approaches to determining the assignment field $u(x,t)$ (2.4) are considered, that establish the correspondence (2.2) of a given feature mapping $g(x,t)$ in two given images.

The following sections conform to a classification of these approaches. Both the scope and the application areas associated with each class of approaches overlap with the variational approach of Sect. 3, but otherwise differ. The presentation focuses on the former aspects and the essential differences, whereas an in-depth discussion of the latter aspects is beyond the scope of this survey.

Section 4.1 discusses local approaches to the assignment problem whereas the remaining three sections are devoted to global approaches. In Sect. 4.2 the correspondence problem is reformulated as an labeling problem so that methods for solving the *Maximum A Posteriori (MAP)* problem with the corresponding *Markov Random Field (MRF)* model can be applied. Assignment by *variational image registration* is briefly considered in Sect. 4.3.

4.1. Local Approaches.

Key feature of the class of assignment approaches is the restriction of the set of feasible assignment fields $u(x)$ to a *finite* set. This set is defined by restricting at each location $\{x^i\}_{i\in[n]} \in \Omega$ the range of $u(x^i) \in \mathcal{U}(x^i)$ to a finite set $\mathcal{U}(x^i)$.

Local approaches determine the optimal $u(x^i)$ independently, i.e. they solve for each $i \in [n]$

$$u(x^i) \in \operatorname*{argmin}_{u\in\mathcal{U}(x^i)} \rho\big(g(x^i,t), g(x^i+u, t+\delta t)\big). \tag{4.1}$$

The usually small sets $|\mathcal{U}(x^i)|$ allow *exhaustive search* to find an optimal solution. Thus, the general distance function $\rho(\cdot,\cdot)$ is not required to be convex or differentiable and allows for more involved formulations.

Since local methods do not make use of (non-local) spatial smoothness priors w.r.t. u, they require – and, in fact, solely rely on – discriminative features, typically derived from local images patches also used by local *variational* methods, see (3.18):

$$g(x^i,t) = \big(f(x^j,t)\big)_{x^j\in\mathcal{N}(x^i)} \in \mathbb{R}^p, \quad p = |\mathcal{N}(x^i)| \tag{4.2}$$

with some neighborhood $\mathcal{N}(x^i)$, e.g. a square region.

In the following some common choices for ρ are addressed. For brevity, the discussion omits references to x^i and some fixed $u = u(x^i)$ and puts $g^1 := g(x^i, t^1)$, $g^2 := g(x^i+u, t^2)$ with $t^2 = t^1 + \delta t$.

Template-based matching methods compare a template g^1 pixel-wise to a potential match g^2 and derives some similarity measure from it. Direct comparison of gray values,

$$\rho(g^1, g^2) = \rho(g^1 - g^2) \tag{4.3}$$

is usually avoided in favor of distance functions which are invariant to brightness or geometric changes. Two popular choices are:

- The *normalized cross-correlation* [114] derives patch features which are invariant to global additive and multiplicative changes of g by defining

$$\overline{g}^k = \frac{g^k - \mu(g^k)}{\sigma(g^k)}, \qquad k = 1,2 \tag{4.4}$$

with mean $\mu(g^k)$ and standard deviation $\sigma(g^k)$ of samples $\{g^k_j\}_{j\in[p]}$. Then the distance function is defined as

$$\rho_{\mathrm{NCC}}(g^1,g^2) = 1 - \frac{1}{p}\langle \overline{g}^1, \overline{g}^2\rangle = \frac{1}{2p}\rho_2^2(\overline{g}^1 - \overline{g}^2) \tag{4.5}$$

where the last equation follows from $\langle \overline{g}^k, \overline{g}^k \rangle = p\sigma^2(\overline{g}^k) = p$.

- The *Census transform* creates binary descriptors

$$\overline{g}^k = \Big(\psi_{\mathbb{R}_+}(g_j^k - m^k)\Big)_{j\in[p]} \in \{0,1\}^p, \qquad k = 1,2 \tag{4.6}$$

with $m^k := g(x^i, t^k)$, which approximate directional derivatives [56] and measures the Hamming distance

$$\rho_{\mathrm{CT}}(g^1, g^2) = \rho_1\left(\overline{g}^1 - \overline{g}^2\right) = \|\overline{g}^1 - \overline{g}^2\|_1 . \tag{4.7}$$

This transformation is in particular invariant to any strictly monotonically increasing transformation $\gamma\colon \mathbb{R} \mapsto \mathbb{R}$ uniformly applied to all components of g^1 and g^2.

Histogram-based methods relax the pixel-by-pixel comparison in (4.3) to achieve additional invariance to geometric transformations.

Exemplarily, a method frequently used in medical images registration [124] and stereo disparity estimation [65] is detailed. It uses the concept of *mutual information* [38] to measure distances between gray-value probability distributions $\hat{\mathrm{p}}_k(f; g^k)$, $k = 1, 2$, determined as kernel density estimates [94] from the samples $\{g_j^k\}_{j\in[p]}$. Their entropies are given by

$$\mathrm{H}(\hat{p}_k; g^k) = -\int \hat{\mathrm{p}}_k(f; g^k)\, \log \hat{\mathrm{p}}_k(f; g^k)\, \mathrm{d}f, \qquad k = 1, 2. \tag{4.8}$$

The joint distribution $\hat{\mathrm{p}}_{1,2}(f^1, f^2; g^1, g^2)$ is defined accordingly with joint entropy

$$\mathrm{H}(\hat{\mathrm{p}}_{1,2}; g^1, g^2) = -\int \hat{\mathrm{p}}_{1,2}(f^1, f^2; g^1, g^2)\, \log \hat{\mathrm{p}}_{1,2}(f^1, f^2; g^1, g^2)\, \mathrm{d}f^1 \mathrm{d}f^2. \tag{4.9}$$

Then the mutual information defines the distance function

$$\rho_{\mathrm{MI}}(g^1, g^2) = \mathrm{H}(\hat{p}_1; g^1) + \mathrm{H}(\hat{p}_2; g^2) - \mathrm{H}(\hat{\mathrm{p}}_{1,2}; g^1, g^2) \tag{4.10}$$

which shows some robustness against rotation, scaling and illumination changes.

Complex approaches such as *Scale-invariant feature transform (SIFT)* [82] and *Speeded Up Robust Features (SURF)* [17] combine several techniques including histogram of orientations and multiple resolution to optimize robustness, reliability and speed.

4.2. **Assignment by Displacement Labeling.** Consider again sets $\mathcal{U}(x^i)$ of assignment vectors as discussed in Sect. 4.1. In constrast to local approaches presented in the previous section, this section is devoted to methods that *simultaneously* select vectors $u(x^i) \in \mathcal{U}(x^i)$ for all locations x^i, $i \in [n]$, based on optimization criteria that evaluate desired properties of assignment fields u. The feasible set of u is denoted by $\mathcal{U} := \cup_{i\in V}\mathcal{U}(x^i)$. It will be convenient to index locations $\{x^i\}_{i\in V} \in \Omega$ by vertices $i \in V = [n]$ of a graph $G = (V, E)$.

As a consequence of the twofold discretization of both the underlying domain $\Omega \subset \mathbb{R}^d$ and the range of $u(x)$, it makes sense to associate with each location x^i an integer-valued variable

$$\ell_i := \ell(x^i) \in [m_i], \qquad m_i := |\mathcal{U}(x^i)|, \tag{4.11}$$

whose value determines the assignment vector $u(x^i) \in \mathcal{U}(x^i)$. This separates the problem formulation in terms of the *labeling field* $\ell := \{\ell_i\}_{i\in V}$ from the set of assignment vectors $\mathcal{U}$ that may vary, as is further discussed below.

Analogous to objectives (3.1) of variational approaches, a functional as criterion for labelings ℓ defines an approach,

$$\begin{aligned} J(\ell; \mathcal{U}) &= J_{\mathrm{D}}(\ell) + J_{\mathrm{R}}(\ell) \\ &= \sum_{i\in V} \varphi_i(\ell_i; \mathcal{U}) + \sum_{ij\in E} \varphi_{ij}(\ell_i, \ell_j; \mathcal{U}), \end{aligned} \tag{4.12}$$

together with an algorithm for determining an assignment field u in terms of a minimizing labeling field ℓ. For instance, in view of a data term like (3.6), a reasonable definition of the function $\varphi_i(\cdot;\mathcal{U})$ of (4.12) is

$$\varphi_i(\ell_i;\mathcal{U}) = \rho_{\mathrm{D}}\left(\left\|\mathrm{J}_g(x^i)u_{\ell_i} + \partial_t g(x^i)\right\|_F\right), \quad u_{\ell_i} \in \mathcal{U}(x^i), \quad \ell_i \in [m_i] \tag{4.13}$$

where ℓ_i enumerates all possible assignment vectors u_{ℓ_i} at x^i. However, getting back to the differences to the differential approach addressed in Sect. 2.2.2, a major motivation of formulation (4.12) is to disregard partial derivatives of the feature map involved in differential variational approaches (Sect. 3.1), and hence to avoid the corresponding limitations discussed in Sections 2.2.3 and 3.3.3. Rather, data terms J_{D} are directly defined by setting up and evaluating *locally* possible assignments $u_{\ell_i} \in \mathcal{U}(x^i)$ that establish a correspondence between local features (2.2), extracted from the given image pair, and by defining costs $\varphi_i(\ell_i;\mathcal{U})$ accordingly. Notice that no smoothness of φ_i is required – any distance discussed in Sect. 4.1 may be employed as in (4.13). For a discussion of the distance (4.10) in this connection, see [75].

The same remarks apply to the definition of J_{R} in (4.12). A common choice in the literature however is the discrete version of the non-convex regularizer (3.42)

$$\varphi_{ij}(\ell_i,\ell_j;\mathcal{U}) = \rho_{2,\lambda}(u_{\ell_i} - u_{\ell_j};\mathcal{U}). \tag{4.14}$$

The reader should notice that the *non-convex* regularizer (3.42) has been replaced by the *combinatorial* version (4.14). Likewise, the *non-convex* data term (3.6) has been replaced by the *discrete-valued* term (4.13). More generally, the problem related to the variational approach to cope with the non-convexity of the data term (Sect. 3.3.3) by means of a multiscale implementation (Sect. 3.3.3), and with the non-convexity of the overall functional by computing a "good" local minimum (Sect. 3.6.3), has been replaced by the combinatorial problem to determine an optimal assignment by minimizing (4.12). This problem is known in the literature as *Maximum A Posteriori (MAP) problem* w.r.t. the discrete probabilistic graphical model

$$\mathrm{p}_G(\ell;\mathcal{U}) = \frac{1}{Z}\exp\big(-J(\ell;\mathcal{U})\big), \qquad Z = \sum_{\ell}\exp\big(-J(\ell;\mathcal{U})\big), \tag{4.15}$$

that is the problem to compute the *mode* $\operatorname{argmax}\mathrm{p}_G(\ell;\mathcal{U})$ of the *Markov Random Field* p_G defined on the undirected graph G. See [127, 77] for background and further details.

Many past and current research activities are devoted to this problem, across various fields of computer science and applied mathematics. Approaches range from integer programming techniques to various convex relaxations and combinations thereof. To get a glimpse of the viewpoint of polyhedral combinatorics on the problem to minimize (4.12), consider a single summand $\varphi_i(\ell_i;\mathcal{U})$ and define the vector

$$\theta^i \in \mathbb{R}^{m_i}, \qquad \theta^i_{\ell_i} := \varphi_i(\ell_i;\mathcal{U}), \qquad \ell_i \in [m_i], \tag{4.16}$$

whose components specify the finite range of the function φ_i. Then the problem of determining $\overline{\ell}_i$ corresponding to the minimal value of $\varphi_i(\ell_i;\mathcal{U})$, can be rewritten as

$$\min_{\mu^i \in \Delta_{m_i}} \langle \theta^i, \mu^i \rangle, \tag{4.17}$$

which is a *linear program (LP)*. Clearly, for general data defining θ^i by (4.16), the vector $\overline{\mu}^i$ minimizing (4.17) is a vertex of the simplex Δ_{m_i} corresponding to the indicator vector $\overline{\mu}^i = (0,\ldots,0,1,0,\ldots,0)^\top$ of the value $\overline{\ell}_i$. This reformulation can be applied in a straightforward way to the overall problem of minimizing (4.12), resulting in the LP

$$\min_{\mu \in \mathcal{M}_G} \langle \theta, \mu \rangle, \tag{4.18}$$

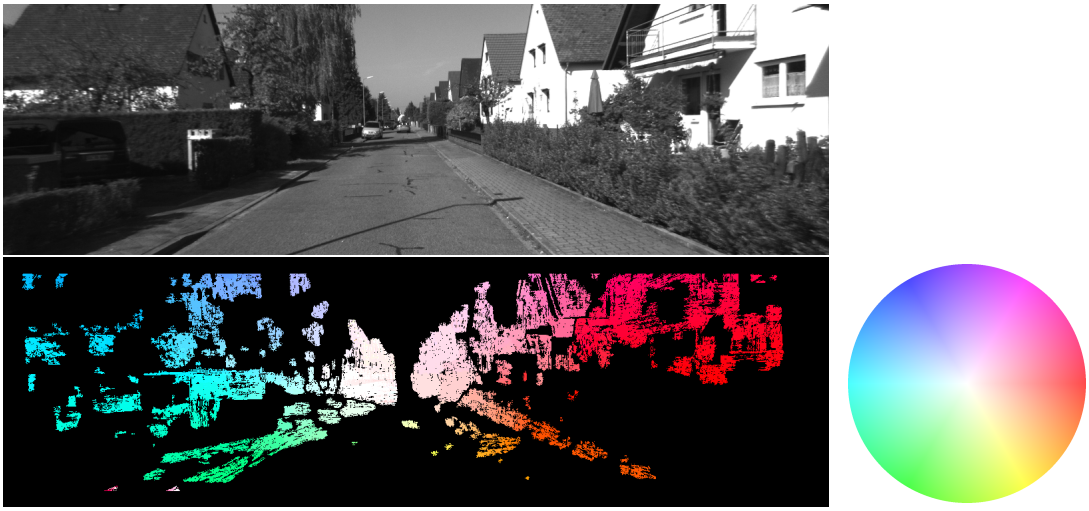


FIGURE 15. TOP Frame of a sequence, taken with a fast moving camera from the KITTI benchmark (Sect. 2.5). BOTTOM Optical flow estimate based on MAP inference. The disk on the right displays the color code of flow vectors. Each image patch localized at x^i where a sufficiently discriminative feature could be extracted, is associated with a set $\mathcal{U}(x^i)$ of possible assignment vectors $u_{\ell_i} \in \mathcal{U}(x^i)$. The displayed assignment field $\overline{u} := \{u_{\overline{\ell}_i}\}_{i \in V}$ is determined by a labeling field $\overline{\ell}$ minimizing the functional (4.12). The latter combinatorial task has been solved to global optimality by an approach combining convex relaxation and integer programming [106]. Global optimality enables model validation: any deficiencies of the assignment field estimate are solely due to the model components, feature extraction and constraints, as encoded by the MRF (4.15) through $J(\ell; \mathcal{U})$.

defined over the so-called *marginal polytope* $\mathcal{M}_G$. This polytope is the convex hull of feasible vectors μ, as is the simplex Δ_{m_i} in (4.17) for the feasible vectors μ^i. The combinatorial complexity of determining the integer-valued minimizer of (4.12) is reflected by the complexity of the marginal polytope $\mathcal{M}_G$. This complexity is due to the interaction of variables ℓ_i, ℓ_j as defined by the edges $ij \in E$ of the underlying graph, besides the integer constraints $\ell_i \in [m_i],\ \forall i \in [n]$.

Formulation (4.18) is the starting point for *convex relaxations* by optimizing over simpler polytopes, defined by a *subset* of inequalities that specify facets of $\mathcal{M}_G$. The recent paper [73] reports a comprehensive evaluation of a broad range of approaches to problem (4.18). Figure 15 illustrates an application to optical flow estimation.

While research on inference methods for graphical models is supporting the design of new approaches to optical flow estimation, the need to restrict the range of u to a finite set $\mathcal{U}$ is a significant restriction. As a consequence, approaches either exploit prior knowledge about u, so as to enable a covering of the relevant range of u with high resolution through the set $\mathcal{U}$ with bounded size $|\mathcal{U}|$, or they solve problem (4.14) once more after refining $\mathcal{U}$, based on a first estimate of u.

For instance, the work [52] exploits the probabilistic model (4.15) in order to estimate locally the uncertainty of a first estimate u, which in turn is used to refine the set $\mathcal{U}$ so as to accommodate the discretization to the local variability of $u(x)$. The approach [90] first determines a coarse estimate of u in a preprocessing stage by global phase-based correlation, followed by defining possible refinements of $u(x)$ in terms of $\mathcal{U}$. The authors of [138] rely on a prior estimate of the fundamental matrix F

(2.34) using standard methods, which enables to properly define $\mathcal{U}$ based on the epipolar constraint (2.38).

In a way, while the former two approaches mimic range refinement of variational methods through representing u at multiple scales (Sect. 3.3.3), the latter approach exploits geometrical prior knowledge in a similar way to variational methods as discussed in Sect. 3.5.2. Future research during the next decade will have to reveal more clearly the pros and cons of these related methods.

4.3. **Variational Image Registration.** The objective of *image registration* is to assign two images in terms of a diffeomorphism $u\colon \Omega \to \Omega$ of the underlying domain. A major motivation for this inherent *smoothness* of u have been applications to computational anatomy [140], based on fundamental work of Grenander, Dupuis, Trouvé, Miller, Younes and others – cf. [42, 19, 140, 139] and references therein.

Another basic motivation for the methodology of image registration is the use of point features, so-called "landmarks", for establishing sparse assignments, that need to be interpolated in a subsequent step to obtain a transform of the entire underlying domain. This is usually accomplished by kernel functions that span a corresponding Hilbert space of smooth functions with bounded point-evaluation functional [136, 31]. Interpolation with thin-plate splines is a well-known example, and extensions to approximating deformations are straightforward. See [98, 88] for corresponding overviews in connection with medical imaging.

The *Large Deformation Diffeomorphic Metric Matching (LDDMM)* approach [19, 50, 139], that emerged from the works cited above, has evolved over the years into a theoretical and computational framework for diffeomorphic image registration. In particular, the application to the assignment of point sets, in connection with kernel functions, leads to a canonical system of ODEs whose numerical solution generates a diffeomorphic assignment along a geodesic path on the diffeomorphism group. See [113] for recent references and an extension for better handling deformations at multiple scales.

The importance of this framework is due to the well-developed mathematical basis, and due to its broad applicability in the fields of computational anatomy and medical imaging. The mathematical relations to continuum and fluid mechanics and the corresponding relevancy to imaging problems with physical prior knowledge (cf. Sect. 3.5.3) are intriguing as well. In the field of computer vision, deformable shape matching constitutes a natural class of applications, unlike the more common optical flow fields in natural videos that typically exhibit discontinuities, caused depth changes and independently moving objects.

5. Open Problems and Perspectives

5.1. **Unifying Aspects: Assignment by Optimal Transport.** The mathematical theory of *optimal transport* [123, 5] provides a general formulation of the assignment problem that bears many relations to the approaches discussed so far.

Consider again the set-up discussed in Sect. 4.2: at each location x^i indexed by vertices $i \in V = [n]$, a vector $u(x^i) \in \mathcal{U}(x^i)$ from a set of candidates $\mathcal{U}(x^i)$ has to be selected. Put $\mathcal{U} = \cup_{i\in[n]}\mathcal{U}(x^i)$. Denote by V' the index set of all locations $\{x^i + u(x^i)\}_{u(x^i)\in\mathcal{U}(x^i)}$, $\forall i \in V$, that u may assign to the locations indexed by V. Then this set-up is represented by the bipartite graph $G = (V, V'; E)$ with edge set $E = \{ij \in V \times V' \colon \exists u \in \mathcal{U},\, x^i + u = x^j\}$. The first term of the objective (4.12) specifies edge weights $\varphi_i(\ell_i;\mathcal{U})$ for each edge corresponding to the assignment $x^i + u_{\ell_i} = x^j$, and minimizing only the first term $\sum_{i\in V}\varphi_i(\ell_i;\mathcal{U})$ would *independently* select a unique vector $u(x^i)$ from each set $\mathcal{U}(x^i)$, $i \in V$, as solution to (4.17).

A classical way to remove this independency is to require the selection of *non-incident* assignments, that is besides uniquely assigning a vector $u \in \mathcal{U}(x^i)$ to x^i, $\forall i \in V$, it is required that there is *at most one* correspondence $x^i + u = x^j$, for all $j \in V'$. This amounts to determining an optimal weighted *matching* in the bipartite graph $G = (V, V'; E)$. Formally, collecting the edge weights $\varphi_i(\ell_i;\mathcal{U})$ by a vector $\theta \in \mathbb{R}^m$, $m = \sum_{i\in V} m_i$, with subvectors given by (4.16), the LP

$$\min_{\mu\in\mathbb{R}^{|E(G)|}} \langle\theta, \mu\rangle \quad \text{subject to} \quad \mu \geq 0, \quad B_G\mu \leq \mathbb{1}_{|V\cup V'|}, \quad B_G \in \{0,1\}^{|V\cup V'|\times|E|}, \tag{5.1}$$

has to be solved where B_G is the incidence matrix of graph G. It is well known that the polyhedron $\mathbb{R}_+^{|E|} \cap \{\mu \colon B_G\mu \leq \mathbb{1}_{|V\cup V'|}\}$ is integral [78], which implies a binary solution $\overline{\mu} \in \{0,1\}^{|E|}$ to (5.1) satisfying the required uniqueness condition. Note that this condition may be regarded as a weak regularity condition enforcing a minimal degree of "smoothness" of the assignment field u.

The connection to optimal transport can be seen by reformulating problem (5.1). Put $n' = |V'|$ and let the matrix $c \in \mathbb{R}^{n\times n'}$ encode the *costs* of assigning (transporting) location x^i to $x^j = x^i + u$, $u \in \mathcal{U}(x^i)$. Then consider the problem

$$\min_{\mu\in\mathbb{R}^{n\times n'}} \langle c, \mu\rangle \text{ subject to } \mu \geq 0,\ \mu\mathbb{1}_{n'} = \mathbb{1}_n,\ \mu^\top\mathbb{1}_n \leq \mathbb{1}_{n'},\ n \leq n', \tag{5.2}$$

where the unknowns are deliberately denoted again by μ. The second constraint says that each node $i \in V$ (location x^i) is uniquely assigned to some node $j \in V'$ (location x^j). The third constraint says that at most one vertex $i \in V$ is assigned to each $j \in V'$. The last condition $n' \geq n$ naturally holds in practical applications. It is straightforward to show [107, Prop. 4.3] that the solution $\overline{\mu} \in \{0,1\}^{n\times n'}$ to (5.2) is again integral.

In the case $n = n'$, problem (5.2) equals the *linear assignment problem*, which is a discrete version of *Monge-Kantorovich* formulation of the *optimal transport* problem. The constraints of (5.2) then define the Birkhoff polytope and the minimizer $\overline{\mu}$ at some vertex of this feasible set is a permutation matrix that uniquely maps V and V' onto each other. Matrices μ that are not vertices (extreme points) of the polytope are doubly stochastic, hence rows $\mu_{i,\bullet} \in \Delta_n$, $i \in [n]$ and columns $\mu_{\bullet,j}$, $j \in [n]$ represent *non-deterministic* assignments of vertices $i \in V$ and $j \in V'$, respectively.

The general formulation [123] considers Polish probability spaces $(\mathcal{X}, \mu_\mathcal{X}), (\mathcal{Y}, \mu_\mathcal{Y})$ with Borel probability measures $\mu_\mathcal{X} \in \mathcal{P}(\mathcal{X}), \mu_\mathcal{Y} \in \mathcal{P}(\mathcal{Y})$, and the set of *coupling measures*, again deliberately denoted by μ, that have $\mu_\mathcal{X}, \mu_\mathcal{Y}$ as marginals,

$$\begin{aligned}\mathcal{M}(\mu_\mathcal{X}, \mu_\mathcal{Y}) = \big\{\mu \in \mathcal{P}(\mathcal{X}\times\mathcal{Y}) \colon \mu(A\times\mathcal{Y}) = \mu_\mathcal{X}(A),\ \mu(\mathcal{X}\times B) = \mu_\mathcal{Y}(B),\\ \forall A \subseteq \mathcal{B}(\mathcal{X}),\ \forall B \subseteq \mathcal{B}(\mathcal{Y})\big\}.\end{aligned} \tag{5.3}$$

Given a Borel cost function $c\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$, the problem analogous to (5.2) in the case $n = n'$ reads

$$\inf_{\mu \in \mathcal{M}(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y) \mathrm{d}\mu(x,y). \tag{5.4}$$

A central question concerns conditions on c that imply existence of *deterministic* minimizers $\overline{\mu}$ of (5.4), that is existence of a measurable function $T\colon \mathcal{X} \to \mathcal{Y}$ such that for random variables (X, Y) with law μ the relation $Y = T(X)$ holds. The *assignment* T is called *transportation map* that "pushes forward" the "mass" represented by $\mu_{\mathcal{X}}$ onto $\mu_{\mathcal{Y}}$, commonly denoted $T_{\#}$:

$$T_{\#}\mu_{\mathcal{X}} = \mu_{\mathcal{Y}} \qquad \text{with} \qquad \mu_{\mathcal{Y}}(B) = \mu_{\mathcal{X}}(T^{-1}(B)), \quad \forall B \in \mathcal{B}(\mathcal{Y}). \tag{5.5}$$

Likewise, $\overline{\mu}$ is concentrated on the graph of T, akin to the concentration of minimizers of (5.2) on a set of binary matrices.

Due to its generality formulation (5.4) provides a single framework for addressing a range of problems, related to optical flow estimation by assignment. This particularly includes:

- The representation of both discrete and continuous settings as sketched above, and the applicability to the assignment of arbitrary objects, as defined by the spaces $\mathcal{X}, \mathcal{Y}$.
- The focus on the combinatorial nature of the assignment problem, on convex duality and tightness or lack of thightness of the convex relaxation (5.4), together with a probabilistic interpretation in the latter case.
- Conservation of mass reflects the invariance assumption underlying (2.2) and (2.6), respectively.
- The differential, dynamic viewpoint: Let $\mathcal{X} = \mathbb{R}^d$ and define the cost function

  $$c(x,y) = \|x - y\|^2 \tag{5.6}$$

  and the *Wasserstein space* $\big(\mathcal{P}_2(\mathcal{X}), W_2\big)$ of Borel probability measures

  $$\mathcal{P}_2(\mathcal{X}) := \big\{\mu_{\mathcal{X}} \in \mathcal{P}(\mathcal{X})\colon \int_{\mathcal{X}} \|x - y\|^2 \mathrm{d}\mu_{\mathcal{X}}(x) < \infty, \ \forall y \in \mathcal{X}\big\}, \tag{5.7}$$

  equipped with the *Wasserstein distance*

  $$W_2(\mu_{\mathcal{X}}, \mu'_{\mathcal{X}}) := \left(\int_{\mathcal{X}\times\mathcal{X}} \|x - y\|^2 \mathrm{d}\overline{\mu}(x,y)\right), \qquad \forall \overline{\mu} \text{ solving (5.4)}, \tag{5.8}$$

  with $\mu_{\mathcal{Y}}$ replaced by $\mu'_{\mathcal{X}}$ in (5.4). Then the path $(\mu_{\mathcal{X},t})$ defined by

  $$\mu_{\mathcal{X},t} = \big((1-t)I + tT\big)_{\#}\mu_{\mathcal{X}} \tag{5.9}$$

  and some optimal map T via (5.5), satisfies the continuity equation

  $$\frac{\mathrm{d}}{\mathrm{d}t}\mu_t + \operatorname{div}(v_t \mu_t) = 0 \tag{5.10}$$

  with velocity field $v_t\colon \mathbb{R}^d \to \mathbb{R}^d$ given by $v_t = (T - I) \circ ((1-t)I + tT)^{-1}$, $\forall t$ in the sense of distributions. Eqn. (5.10) provides a natural connection to continuum and fluid mechanics (cf. e.g. [22]) and also to flows generating diffeomorphic assignments under additional assumptions [139, Ch. 11]. Comparing (5.10) and (3.3) shows that, if g is regarded as a density for the scalar case $p = 1$, condition (3.2) is a strong assumption implying $\operatorname{div} u = 0$.

The generality of this framework explains too, however, why the regularity of solutions to the Monge-Kantorovich problem is a subtle issue, even when given as deterministic assignment T. This is also apparent through Euler's equation (5.10), which lacks any viscous term that would induce some regularity.

From this viewpoint, much of the research related to variational optical flow estimation, and to the related problems discussed in Sect. 4, can be understood as

(i) interplay between modelling additional terms that induce a desired degree of spatial regularity, and
(ii) investigation of how this affects relaxation of the assignment problem from the optimization point of view, and the accuracy of its solution.

As a consequence, *no sharp boundaries can* (and should) *be defined that separate these subfields of research*. For instance,

- the paper [21] suggested an early heuristic attempt to combine bipartite graph matching and thin-plate spline based registration.
- The work [51] combines smoothing with radial basis functions and MRF-based labeling (Sect. 4.2) for medical image registration.
- More generally, concerning image labeling, modelling spatial context by the edge-indexed terms φ_{ij} of the objective (4.12) entails the need to relax combinatorially complex polyhedral feasible sets like the marginal polytope in (4.18), whose vertices may *not* correspond to deterministic assignments, unlike assignments as solutions in the simpler case (5.1).
- The authors of [7] introduce a smoothing operator to solve numerically the Monge-Kantorovich problem.
- In [137] a related objective from continuum mechanics is proposed that, for a limiting value of some parameter, models a *viscous* fluid, hence ensures spatial regularity in a physically plausible way, as opposed to the pure continuity equation (5.10) that is lacking any such term. Assignments are computed by numerically tracing corresponding geodesic paths.
- Much more general objectives for assignments are addressed in [87] that take explicitly into account the metric structure of the underlying space $\mathcal{X}$. The problem to "linearize" this combinatorially complex objective in terms of the Monge-Kantorovich problem is studied in [107], along with the problem to define a cost function c so as to preserve the discriminative power of the original objective as much as possible.
- The recent work [108] exploits the Wasserstein distance (5.8) so as to solve simultaneously template-based assignment and image segmentation, by globally minimizing a corresponding joint variational objective.

This sample of the literature suggests to conclude that in the field of variational image registration (e.g. [139, 137]), sophisticated variational approaches exist that are satisfying in both respects (i),(ii) discussed above: These approaches clearly exhibit their properties mathematically, and they induce regularity without compromising accuracy of assignments, due to a good agreement with the physical properties of the objects being matched.

Outside these fields, a similar quality only holds for variational approaches to optical flow estimation that are constrained by – again: physically motivated – state equations (Sect. 3.5.3). A similar level of rigour has not been reached yet in a major application area of optical flow estimation: motion based analysis of videos of unrestricted scenes with uncontrolled viewpoint changes, and with independently moving rigid and articulated objects. This deficiency of related models is aggravated by the need for natural extensions of frame-to-frame assignments to the permanent analysis of *dynamic* scenarios over time (cf. Sect. 5.3).

5.2. **Motion Segmentation, Compressive Sensing.** Research on *compressive sensing* [33, 40] and corresponding applications have been pervading all fields of empirical data analysis, including image reconstruction and more recently video analysis. A central theme are provable guarantees of signal

recovery in polynomial runtime using sub-Nyquist sampling rates and convex relaxations of combinatorial objective functions for signal reconstruction. For instance, the most common scenario concerns the recovery of $u \in \mathbb{R}^n$ from $m \ll n$ linear measurements $Au = b \in \mathbb{R}^m$, by minimizing

$$\min_u \|u\|_1 \qquad \text{subject to} \qquad Au = b, \tag{5.11}$$

under the assumption that u is k-sparse, i.e.

$$\|u\|_0 := |\operatorname{supp}(u)| = \big|\{i \in [n] \colon u_i \neq 0\}\big| \leq k. \tag{5.12}$$

The objective in (5.11) constitutes a convex relaxation of the combinatorial objective $\|u\|_0$ and suitable conditions on A, e.g. A is close to an isometry on the subset of $2k$-sparse vectors, guarantees unique recovery of u with high probability.

This section presents next an extension of this basic reconstruction principle to video analysis by sketching the recent work reported by [58]. Let

$$f_t \in \mathbb{R}^n, \qquad t \in [T], \tag{5.13}$$

denote the raw image sequence data in terms of vectorized image frames f_t, $t = 1, 2, \ldots, T$. Assuming a stationary camera as in surveillance applications, the objective is to separate the static background from objects moving in the foreground. The ansatz is based on the following modelling assumptions:

- At each point of time $t \in T$, image data are only sampled on a subset $\Omega_t \subset \Omega$ of the discretized domain Ω, resulting in subvectors
$$f_{\Omega_t}, \qquad t \in [T]. \tag{5.14}$$
The sample set Ω_t may vary with t.
- The variation of components of f_{Ω_t} corresponding to the static background is caused by global smooth illumination changes. Hence, this part of f_{Ω_t} can be represented by a low-dimensional subspace
$$U_{\Omega_t} v_t, \qquad U_{\Omega_t} \in \mathbb{R}^{|\Omega_t| \times n_U}, \qquad t \in [T], \tag{5.15}$$
generated by n_U orthonormal colums of a matrix U_t that are subsampled on Ω_t, and some coefficient vector v_t. Research in computer vision [20, 16] supports this subspace assumption.
- Objects moving in the foreground cover only small regions within Ω. Hence they can be represented by vectors
$$s_{\Omega_t} \qquad \text{with} \qquad |\operatorname{supp}(s)| \ll n. \tag{5.16}$$

Putting all together the model reads

$$f_{\Omega_t} = U_{\Omega_t} v_t + s_{\Omega_t}, \qquad t \in [T], \tag{5.17}$$

and convex relaxation of minimizing $|\operatorname{supp}(s)|$ due to (5.16) leads to the recovery approach

$$\min_{U, v_t, s_{\Omega_t}} \|s_{\Omega_t}\|_1 \qquad \text{subject to} \qquad U_{\Omega_t} v_t + s_{\Omega_t} = f_{\Omega_t}. \tag{5.18}$$

Comparison to (5.11) shows similar usage of the sparsity-inducing ℓ^1 norm and subsampled measurements (5.14) as input data. On the other hand, the low-dimensional representation (5.15) of the static part of the video is estimated as well, and the entire video is recovered in terms of U_t (hence U rather than U_{Ω_t} is optimized in (5.18)). In fact, this joint optimization problem is *non*-convex and handled in [58] by alternating optimization:

– For *fixed* U_t, problem (5.18) is solved by applying ADMM (cf. Sect. 3.6.2) to the augmented Lagrangian $L_\lambda(U, v_t, s_{\Omega_t}, w_{\Omega_t})$ with multiplier vector w_{Ω_t} and parameter λ as in (3.40).

– Having determined $v_t, s_{\Omega_t}, w_{\Omega_t}$, the subspace U_t is tracked by performing gradient descent with respect to $L(\cdot, v_t, s_{\Omega_t}, w_{\Omega_t})$ on the Grassmannian $\mathcal{G}(n_U, \mathbb{R}^n)$ (cf., e.g. [1]), resulting in U_{t+1}.

The closely related static viewpoint on the same problem reveals its relevancy to several important research directions. Let

$$F = [f_1, \dots, f_T] = L + S \tag{5.19}$$

denote the whole video data that, due to the reasoning above, are supposed to be decomposable into a *low-rank* matrix L and a *sparse* matrix S. The corresponding convex relaxation approach [34] reads

$$\min_{L,S} \ \|L\|_* + \alpha\|S\|_1 \qquad \text{subject to} \qquad L + S = F, \tag{5.20}$$

where $\|L\|_* = \sum_i \sigma_i(L)$ denotes the nuclear norm in terms of the singular values of L and $\|S\|_1 = \sum_{i,j} |S_{ij}|$. Here, the nuclear norm $\|\cdot\|_*$ constitutes a convex relaxation of the combinatorial task to minimize the rank of L, analogous to replacing the combinatorial objective $\|u\|_0$ in (5.12) by $\|u\|_1$ in (5.11). Clearly, the online ansatz (5.17) along with the corresponding incremental estimation approach is more natural for processing *long* videos. The price to pay is the need to cope with a non-convex (albeit smooth) problem, whereas the batch approach (5.20) is convex.

Future research will tackle the challenging, more general case of non-static backgrounds and moving cameras, respectively. For scenarios with small displacements $u(x)$, work that represents the state of the art is reported in [11]. Results in computer vision that support subspace models and low-rank assumptions have been established [70], and the problem of clustering data lying in unknown low-dimensional subspaces has received considerable attention [80, 4, 43]

From a broader perspective, video analysis and motion-based segmentation provides attractive connections to research devoted to union-of-subspaces models of empirical data and relevant compressive sensing principles [83, 35, 96], and to advanced *probabilistic* models and methods for nonparametric inference [120, 66].

5.3. **Probabilistic Modelling and Online Estimation.** There is a need for advanced probabilistic models and three related aspects of increasing difficulty are briefly addressed:

- A persistent issue of most variational models of mathematical imaging, including those for optical flow estimation, concerns the *selection of* appropriate *hyperparameter values*, like the parameter σ of (3.7) weighting the combination of data term and regularizer (3.1). In principle, Bayesian hierarchical modelling [39] provides the proper framework for calibrating variational models in this respect. The paper [59] illustrates an application in connection with optical flow estimation, based on the marginal data likelihood [85] interpreted as hyperparameter (model) evidence.

  Estimating hyperparameter values from given data in this way entails the evaluation of high-dimensional integrals for marginalization, commonly done using Laplace's method and a corresponding approximation by Gaussian (quadratic) integrals [117, 74]. A validation for complex high-dimensional posterior distributions encountered in variational imaging is involved, however, and is also stimulating more recent research in the field of statistics [99].

  Using discrete variational models (Sect. 4.2) aggravates this problem, due to considerable computational costs and since no widely accepted methods have been established analogous to the above-mentioned approximations.
- Computational costs in connection with runtime requirements become a serious problem when *dynamic scenarios* are considered. While extensions of the domain to $\Omega \times [0, T]$ like in (3.27) are straightforward mathematically and have proven to significantly increase accuracy of optical flow estimation, employing a static model in terms of elliptic Euler-Lagrange systems to

a dynamic system appears somewhat odd, not to mention the need to shift the time interval $[0, T]$ along the time axis in order to analyze long image sequences.

Such extensions appear more natural in connection with dynamic physical models constraining optical flow estimation, as opposed to stationary formulations like (3.33). See [9] for a corresponding approach to data assimilation [116]. A nice feature of this method is the ability to estimate initial conditions that are generally unknown, too. On the other hand, the computational costs necessitate to propagate a low-dimensional POD-projection of the state variables (POD: proper orthogonal decomposition) since the control of dynamical systems [53] entails looping forward and backward through the entire time interval.

- The last remark points to the need for *online estimation methods* that are *causal and optimal*, in connection with the analysis of dynamical system through image analysis. Again the proper framework is known since decades: Given stochastic state and observation processes

$$S = \{S_t\}_{t\geq 0}, \qquad G = \{G_t\}_{t\geq 0}, \tag{5.21}$$

stochastic filtering [13] amounts to determine the conditional distribution of S_t given the observation history and to evaluate it in terms of expectations of the form $\mathbb{E}[\varphi(S_t)|g_s,\, 0 \leq s \leq t]$, for some statistic $\varphi(\cdot)$ of interest (e.g. simply $\varphi(S_t) = S_t$) and conditioned on realizations g_s of G_s, $s \in [0, t]$. Most research during the last decade considered the design of particle filters [41, 13] to the estimation of *low*-dimensional states based on image measurements. This does not scale-up however to high-dimensional states like optical flows $S_t = u_t$.

An attempt to mimic online estimation in connection with instationary optical flows related to experimental fluid dynamics is presented in [103], with states and their evolution given by vorticity transport. For low signal-to-noise ratios and sufficiently high frame rates, the approach performs remarkably well. Another dynamical computer vision scenario is discussed in the recent work [18]. Here the states $S_t = (z_t, \{h_t, R_t\}) \in \mathbb{R}^n \times \mathsf{SE}(3)$ are dense depth-maps z_t (cf. (3.28)) together with varying motion parameters $\{h_t, R_t\}$ describing the observer's motion relative to the scene, to be estimated from image sequence features g_t as measurements via optical flow estimates u_t – see Fig. 13. The approach involves prediction and fusion steps based on Gaussian approximation and joint optimization, yet cannot be considered as direct application of the stochastic filtering framework, in a strict sense. This assessment applies also to labeling approaches (Sect. 4.2) and their application to dynamic scenarios.

6. Conclusion

Optical flow estimates form an essential basis for low-level and high-level image sequence analysis and thus are relevant to a wide range of applications. Corresponding key problems, concepts and their relationships were presented, along with numerous references to the literature for further study. Despite three decades of research, however, an overall coherent framework that enables to mathematically model, predict and estimate the performance of corresponding computational systems in general scenarios, is still lacking. This short survey will hopefully stimulate corresponding methodological research.

APPENDIX A. BASIC NOTATION

Table 1: List of major symbols used in the text.

Symbol	Brief Description	Reference
r.h.s.	abbr.: right-hand side (of some equation)	
w.r.t.	abbr.: with respect to	
w.l.o.g.	without loss of generality	
LP	linear program	
$\mathbb{1}_n \in \mathbb{R}^n$	$(1, 1, \dots, 1)^\top$	
$[n]$, $n \in \mathbb{N}$	integer range $\{1, 2, \dots, n\}$	
$[n]_0$, $n \in \mathbb{N}$	integer range $\{0, 1, \dots, n-1\}$	
$\Omega \subset \mathbb{R}^d$	image domain; typically $d \in \{2, 3\}$	
$x = \big(x_1, \dots, x_d\big)^\top \in \Omega$	image point	
$u(x,t) \in \mathbb{R}^d$	assignment, motion or optical flow field	(2.4), (2.7), (2.8) Sections 2.3.2, 2.3.3
$X = (X_1, X_2, X_3)^\top \in \mathbb{R}^3$	scene point	
$y \in \mathbb{P}^2$, $Y \in \mathbb{P}^3$	homogeneous representation of image and scene points x and X, resp.	Section 2.3.1
$\mathsf{SO}(3)$, $\mathfrak{so}(3)$	special orthog. group and its Lie algebra	
$\mathsf{SE}(3)$	group of Euclidean (rigid) transf. of $\mathbb{R}^3$	
$\{h, R\} \in \mathsf{SE}(3)$	Euclidean transformation of $\mathbb{R}^3$	(2.28)
$[q]_\times \in \mathfrak{so}(3)$, $q \in \mathbb{R}^3$	skew-symm. matrix defined by $[q]_\times X = q \times X$, $\forall X \in \mathbb{R}^3$	
$K \in \mathbb{R}^{3\times 3}$	camera matrix (internal parameters)	Section 2.3.1
$F, E \in \mathbb{R}^{3\times 3}$	fundamental and essential matrix	Section 2.3.1
$f(x,t)$, $x \in \Omega$, $t \in \mathbb{R}$	image sequence	
$\partial_i = \frac{\partial}{\partial x_i}$, $i \in [d]$	spatial partial derivative	
$\partial_t = \frac{\partial}{\partial t}$	temporal partial derivative	
$\partial^\alpha = \frac{\partial^{\lvert\alpha\rvert}}{\partial_1^{\alpha_1} \dots \partial_d^{\alpha_d}}$	multi-index notation	
$\alpha \in \mathbb{N}^d$, $\lvert\alpha\rvert = \sum_{i\in[d]} \alpha_i$		
$\omega^\alpha = \omega_1^{\alpha_1} \cdots \omega_d^{\alpha_d}$	monomial from $\omega \in \mathbb{R}^d$	
$\nabla f(x,t) = \begin{pmatrix} \partial_1 f(x,t) \\ \dots \\ \partial_d f(x,t) \end{pmatrix}$	spatial gradient	
$\nabla_t f(x,t) = \begin{pmatrix} \nabla f(x,t) \\ \partial_t f(x,t) \end{pmatrix}$	spatio-temporal gradient	
$\operatorname{div} u$	divergence $\sum_{i\in[d]} \partial_i u_i$ of a vector field u	
Δ	Laplace operator $\sum_{i\in[d]} \partial_i^2$	
$g(x,t) \in \mathbb{R}^p$, $p \geq 1$	feature mapping (specific meaning and p depend on the context)	
$\mathrm{J}_g(x) = \Big(\big(\nabla g_i(x)\big)_j\Big)_{i\in[p], j\in[d]}$	Jacobian matrix of $g(x) \in \mathbb{R}^p$ at $x \in \mathbb{R}^d$	
$\mathrm{J}_{g,t}(x,t) = \Big(\big(\nabla_t g_i(x,t)\big)_j\Big)_{\substack{i\in[p] \\ j\in[d]\cup\{t\}}}$	Jacobian of $g(x,t) \in \mathbb{R}^p$ at $(x,t) \in \mathbb{R}^{d+1}$	
$\hat{g}(\omega) = \mathcal{F}g(\omega) = \big(\mathcal{F}g\big)(\omega)$	Fourier transform of g	page 8, (2.17)

Continued on next page

Table 1 – continued from previous page

Symbol	Brief Description	Reference
$\langle x, x' \rangle = \sum_i x_i x'_i$	Euclidean inner product	
$\|x\| = \langle x, x \rangle^{1/2}$	Euclidean ℓ^2 norm	
$\|x\|_1 = \sum_i \|x_i\|$	ℓ^1 norm	
$\operatorname{diag}(x)$	diagonal matrix with vector x as diagonal	
$\ker A$	nullspace of the linear mapping A	
$\operatorname{tr} A = \sum_i A_{i,i}$	trace of matrix A	
$\langle A, B \rangle = \operatorname{tr}(A^\top B)$	matrix inner product	
$\|A\|_F = \langle A, A \rangle^{1/2}$	Frobenius norm	
$\rho \colon \mathbb{R} \to \mathbb{R}_+$	distance function	page 6, (2.3)
$\delta_C(x) = \begin{cases} 0, & x \in C \\ +\infty, & x \notin C \end{cases}$	indicator function of a closed convex set $C \subseteq \mathbb{R}^d$	
Π_C	orthogonal projection onto a closed convex set C	
$\Delta_n \subset \mathbb{R}^n$	probability simplex $\{x \in \mathbb{R}^n \colon \sum_{i \in [n]} x_i = 1;\ x \geq 0\}$	page 34, (4.16)

APPENDIX B. CROSS-REFERENCES

- Compressive Sensing
- Duality and Convex Programming
- Energy Minimization Methods
- Graph Cuts
- Iterative Solution Methods
- Large-Scale Inverse Problems in Imaging
- Linear Inverse Problems
- Registration
- Regularization Methods for Ill-Posed Problems
- Splines and Multiresolution Analysis
- Statistical Methods in Imaging
- Total Variation in Imaging

REFERENCES

1. P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, 2008.
2. E.H. Adelson and J.R. Bergen, *Spatiotemporal Energy Models for the Perception of Motion*, J. Opt. Soc. Am. A **2** (1985), no. 2, 284–299.
3. R. J. Adrian and J. Westerweel, *Particle Image Velocimetry*, Cambridge University Press, 2011.
4. A. Aldroubi, *The subspace segmentation problem, nonlinear approximations and applications*, ISRN Signal Proc. **Art.** (2013), no. 417492, 13 pages.
5. L. Ambrosio and N. Gigli, *A User's Guide to Optimal Transport*, Modelling and Optimisation of Flows on Networks, Lect. Not. Math., vol. 2062, Springer, 2013, pp. 1–155.
6. P. Anandan, *A Computational Framework and an Algorithm for the Measurement of Visual Motion*, Int. J. Comp. Vision **2** (1989), 283–310.
7. S. Angenent, S. Haker, and A. Tannenbaum, *Minimizing Flows for the Monge–Kantorovich Problem*, SIAM J. Math. Anal. **35** (2003), no. 1, 61–97.
8. E. Arnaud, E. Mémin, R. Sosa, and G. Artana, *A Fluid Motion Estimator for Schlieren Image Velocimetry*, Proc. ECCV (A. Leonardis, H. Bischof, and A. Pinz, eds.), LNCS, vol. 3951, Springer, 2006, pp. 198–210.
9. G. Artana, A. Cammilleri, J. Carlier, and E. Mémin, *Strong and Weak Constraint Variational Assimilations for Reduced Order Fluid Flow Modeling*, J. Comp. Physics **231** (2012), no. 8, 3264–3288.
10. A. Auslender and M. Teboulle, *Asymptotic Cones and Functions in Optimization and Variational Inequalities*, Springer, 2003.
11. A. Ayvaci, M. Raptis, and S. Soatto, *Sparse Occlusion Detection with Optical Flows*, Int. J. Comp. Vision **97** (2012), no. 3, 322–338.
12. F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, *Optimization with Sparsity-Inducing Penalties*, Found. Trends Mach. Learning **4** (2012), no. 1, 1–106.
13. A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering*, Springer, 2009.
14. S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, *A Database and Evaluation Methodology for Optical Flow*, Int. J. Comput. Vision **92** (2011), no. 1, 1–31.
15. J. L. Barron, D. Fleet, and S. S. Beauchemin, *Performance of Optical Flow Techniques*, Int. J. Comp. Vision **12** (1994), no. 1, 43–77.
16. R. Basri and D.W. Jacobs, *Lambertian Reflectance and Linear Subspaces*, IEEE Trans. Patt. Anal. Mach. Intell. **25** (2003), no. 2, 218–233.
17. H. Bay, A. Ess, T. Tuytelaars, and L. Ban Gool, *Speeded-Up Robust Features (SURF)*, Comp. Vision Image Underst. **110** (2008), no. 3, 346–359.
18. F. Becker, F. Lenzen, J. H. Kappes, and C. Schnörr, *Variational Recursive Joint Estimation of Dense Scene Structure and Camera Motion from Monocular High Speed Traffic Sequences*, Int. J. Comput. Vision **105** (2013), 269–297 (English).
19. M.F. Beg, M.I. Miller, A. Trouv'e, and L. Younes, *Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms*, Int. J. Comp. Vision **61** (2005), no. 2, 139–157.
20. P.N. Belhumeur and D.J. Kriegman, *What Is the Set of Images of an Object Under All Possible Illumination Conditions?*, Int. J. Comp. Vision **28** (1998), no. 3, 245–260.
21. S. Belongie, J. Malik, and J. Puzicha, *Shape Matching and Object Recognition Using Shape Contexts*, IEEE Trans. Patt. Analysis Mach. Intell. **24** (2002), no. 24, 509–522.
22. J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math. **84** (2000), no. 3, 375–393.
23. C.P. Bernard, *Discrete Wavelet Analysis for Fast Optic Flow Computation*, Appl. Comput. Harmonic Anal. **11** (2001), 32–63.
24. M.J. Black and P. Anandan, *The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields*, Comp. Vision Image Underst. **63** (1996), no. 1, 75–104.
25. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, Found. Trends Mach. Learning **3** (2010), no. 1, 1–122.
26. W.L. Briggs, V.E. Henson, and S.F. McCormick, *A Multigrid Tutorial*, 2nd ed., SIAM, 2000.
27. M.Z. Brown, D. Burschka, and G.D. Hager, *Advances in Computational Stereo*, IEEE Trans. Patt. Anal. Mach. Intell. **25** (2003), no. 8, 993–1008.
28. A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger, and C. Schnörr, *Variational optic flow computation in real-time*, IEEE Trans. Image Proc. **14** (2005), no. 5, 608–615.

29. A. Bruhn, J. Weickert, T. Kohlberger, and C. Schnörr, *A Multigrid Platform for Real-Time Motion Computation with Discontinuity-Preserving Variational Methods*, Int. J. Computer Vision **70** (2006), no. 3, 257–277.
30. A. Bruhn, J. Weickert, and C. Schnörr, *Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods*, Int. J. Comp. Vision **61** (2005), no. 3, 211–231.
31. M.D. Buhmann, *Radial Basis Functions*, Cambridge Univ. Press, 2003.
32. D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, *A Naturalistic Open Source Movie for Optical Flow Evaluation*, Proc. ECCV (A. Fitzgibbon et al. (Eds.), ed.), Part IV, LNCS 7577, Springer, 2012, pp. 611–625.
33. E. Candès, J. Romberg, and Tao T., *Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information*, IEEE Trans on Information Theory **52** (2006), 489–509.
34. E.J. Candés, X. Li, Y. Ma, and J. Wright, *Robust Principal Component Analysis?*, J. ACM **58** (2011), no. 3, Article no. 11.
35. L. Carin, R.G. Baraniuk, V. Cevher, V. Dunson, M.I. Jordan, G. Sapiro, and M.B. Wakin, *Learning Low-Dimensional Signal Models*, IEEE Signal Proc. Mag. **28** (2011), no. 2, 39–51.
36. A. Chambolle and T. Pock, *A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging*, J. Math. Imaging Vis. **40** (2011), no. 1, 120–145.
37. P.L. Combettes and J.-C. Pesquet, *Proximal Splitting Methods in Signal Processing*, Fixed-Point Algorithms for Inverse Problems in Science and Engineering (H.H. Bauschke, R. Burachik, P.L. Combettes, V. Elser, D.R. Luke, and H. Wolkowicz, eds.), Springer, New York, 2010.
38. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.
39. N. Cressie and C.E. Wikle, *Statistics for Spatio-Temporal Data*, Wiley, 2011.
40. D. Donoho, *Compressed Sensing*, IEEE Trans. Information Theory **52** (2006), 1289–1306.
41. A. Doucet, S. Godsil, and C. Andrieu, *On Sequential Monte Carlo Sampling Methods for Bayesian Filtering*, Statist. Computing **10** (2000), 197–208.
42. P. Dupuis, U. Grenander, and M.I. Miller, *Variational Problems on Flows of Diffeomorphisms for Image Matching*, Quart. Appl. Math. **56** (1998), no. 3, 587–600.
43. E. Elhamifar and R. Vidal, *Sparse Subspace Clustering: Algorithm, Theory, and Applications*, IEEE Trans. Pattern Anal. Mach. Intell. **35** (2013), no. 11, 2765–2781.
44. A. Elmoataz, O. Lezoray, and S. Bougleux, *Nonlocal Discrete Regularization on Weighted Graphs: A Framework for Image and Manifold Processing*, IEEE Trans. Image Proc. **17** (2008), no. 7, 1047–1059.
45. O. Faugeras and Q.-T. Luong, *The Geometry of Multiple Images*, MIT Press, 2001.
46. D.J. Fleet and A.D. Jepson, *Computation of Component Image Velocity from Local Phase Information*, Int. J. Comp. Vision **5** (1990), no. 1, 77–104.
47. A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, *Vision meets Robotics: The KITTI Dataset*, International Journal of Robotics Research (IJRR) **32** (2013), no. 11, 1231–1237.
48. G. Gilboa and S. Osher, *Nonlocal Operators with Applications to Image Processing*, Multiscale Model. Simul. **7** (2008), no. 3, 1005–1028.
49. V. Girault and P.-A. Raviart, *Finite Element Methods for Navier-Stokes Equations*, Springer, 1986.
50. J. Glaunès, A. Qiu, M.I. Miller, and L. Younes, *Large Deformation Diffeomorphic Metric Curve Mapping*, Int. J. Comp. Vision **80** (2008), no. 3, 317–336.
51. B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios, *Dense Image Registration through MRFs and Efficient Linear Programming*, Medical Image Analysis **12** (2008), 731–741.
52. B. Glocker, N. Paragios, N. Komodakis, G. Tziritas, and N. Navab, *Optical Flow Estimation with Uncertainties through Dynamic MRFs*, Proc. CVPR, 2008.
53. M.D. Gunzburger, *Perspectives in Flow Control and Optimization*, SIAM, 2003.
54. M.D. Gunzburger and S. Manservisi, *Analysis and Approximation of the Velocity Tracking Problem for Navier-Stokes Flows with Distributed Control*, SIAM J. Numer. Anal. **37** (2000), no. 5, 1481–1512.
55. P. Gwosdek, A. Bruhn, and J. Weickert, *Variational Optic Flow on the Sony Playstation 3 – Accurate Dense Flow Fields for Real-time Applications*, J. Real-Time Imag. Proc. **5** (2010), no. 3, 163–177.
56. D. Hafner, O. Demetz, and J. Weickert, *Why Is the Census Transform Good for Robust Optic Flow Computation?*, Proc. SSVM (A. Kuijper, K. Bredies, T. Pock, and H. Bischof, eds.), LNCS, vol. 7893, Springer, 2013, pp. 210–221.
57. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge Univ. Press, 2000.
58. J. He, L. Balzano, and A. Szlam, *Incremental Gradient on the Grassmannian for Online Foreground and Background Separation in Subsampled Video*, Proc. CVPR, 2012.
59. P. Héas, C. Herzet, and E. Mémin, *Bayesian Inference of Models and Hyperparameters for Robust Optical-Flow Estimation*, IEEE Trans. Image Proc. **21** (2012), no. 4, 1437–1451.
60. D.J. Heeger, *Optical Flow Using Spatiotemporal Filters*, Int. J. Comp. Vision **1** (1988), no. 4, 279–302.

61. D. Heitz, E. Mémin, and C. Schnörr, *Variational Fluid Flow Measurements from Image Sequences: Synopsis and Perspectives*, Exp. Fluids **48** (2010), no. 3, 369–393.
62. F. Heitz and P. Bouthemy, *Multimodal Estimation of Discontinuous Optical Flow Using Markov Random Fields*, IEEE Trans. Patt. Anal. Mach. Intell. **15** (1993), no. 12, 1217–1231.
63. E.C. Hildreth, *The Computation of the Velocity Field*, Proc. Royal Soc. B **221** (1984), 189–220.
64. W. Hinterberger, O. Scherzer, C. Schnörr, and J. Weickert, *Analysis of Optical Flow Models in the Framework of Calculus of Variations*, Numer. Funct. Anal. Optimiz. **23** (2002), no. 1/2, 69–89.
65. H. Hirschmuller, *Stereo Processing by Semiglobal Matching and Mutual Information*, IEEE Trans. Patt. Anal. Mach. Intell. **30** (2008), no. 2, 328–341.
66. N.L. Hjort, C. Holmes, P. Müller, and S.G. Walker (eds.), *Bayesian Nonparametrics*, Cambridge Univ. Press, 2010.
67. L.T. Hoai An and T. Pham Dinh, *The DC (Difference of Convex Functions) Programming and DCA Revisited with DC Models of Real World Nonconvex Optimization Problems*, Annals of Operations Res. **133** (2005), 23–46.
68. B.K.P. Horn and B.G. Schunck, *Determining Optical Flow*, Artif. Intelligence **17** (1981), no. 1-3, 185–203.
69. R. Horst and N.V. Thoai, *DC Programming: Overview*, J. Optimiz. Theory Appl. **103** (1999), no. 1, 1–43.
70. M. Irani, *Multi-Frame Correspondence Estimation Using Subspace Constraints*, Int. J. Comp. Vision **48** (2002), no. 3, 173–194.
71. K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, Advances in Design and Control, vol. 15, SIAM, 2008.
72. K. Kanatani, *Transformation of Optical Flow by Camera Rotation*, IEEE Trans. Patt. Anal. Mach. Intell. **10** (1988), no. 2, 131–143.
73. J. Kappes, B. Andres, F. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B. Kausler, J. Lellmann, N. Komodakis, and C. Rother, *A Comparative Study of Modern Inference Techniques for Discrete Energy Minimization Problem*, Proc. CVPR, 2013.
74. R.E. Kass, L. Tierney, and J.B. Kadane, *The Validity of Posterior Expansions Based on Laplace's Method*, Bayesian and Likelihood Methods in Statistics and Econometrics, Elsevier Sci. Publ. B.V., 1990, pp. 473–488.
75. J. Kim, V. Kolmogorov, and R. Zabih, *Visual Correspondence Using Energy Minimization and Mutual Information*, Proc. ICCV, 2003.
76. S. Kindermann, S. Osher, and P.W. Jones, *Deblurring and Denoising of Images by Nonlocal Functionals*, Multiscale Model. Simul. **4** (2005), no. 4, 1091–1115.
77. D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, MIT Press, 2009.
78. B. Korte and J. Vygen, *Combinatorial Optimization*, 4th ed., Springer, 2008.
79. P. Krähenbühl and V. Koltun, *Efficient Nonlocal Regularization for Optical Flow*, Proc. ECCV (A. W. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, eds.), LNCS, vol. 7572, Springer, 2012, pp. 356–369.
80. G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, *Robust Recovery of Subspace Structures by Low-Rank Representation*, IEEE Trans. Patt. Anal. Mach. Intell. **35** (2013), no. 1, 171–184.
81. H.C. Longuet-Higgins and K. Prazdny, *The Interpretation of a Moving Retinal Image*, Proc. Royal Soc. London B **208** (1980), 385–397.
82. D.G. Lowe, *Distinctive Image Features from Scale-Invariant Keypoints*, Int. J. Comp. Vision **60** (2004), no. 2, 91–110.
83. Y.M. Lu and M.N. Do, *A Theory for Samping Signals From a Union of Subspaces*, IEEE Trans. Signal Proc. **56** (2008), no. 6, 2334–2345.
84. B.D. Lucas and T. Kanade, *An Iterative Image Registration Technique with an Application to Stereo Vision*, Proc. IJCAI, vol. 2, 1981, pp. 674–679.
85. D.J.C. MacKay, *Bayesian Interpolation*, Neural Comp. **4** (1992), no. 3, 415–447.
86. J. Margarey and N. Kingsbury, *Motion Estimation Using a Complex-Valued Wavelet Transform*, IEEE Trans. Signal Proc. **46** (1998), no. 4, 1069–1084.
87. F. Mémoli, *Gromov-Wasserstein Distances and the Metric Approach to Object Matching*, Found. Comp. Math. **11** (2011), 417–487.
88. J. Modersitzki, *Numerical Methods for Image Registration*, Oxford Univ. Press, 2004.
89. J.-M. Morel and G. Yu, *ASIFT: A New Framework for Fully Affine Invariant Image Comparison*, SIAM J. Imag. Sci. **2** (2009), no. 2, 438–469.
90. M.G. Mozerov, *Constrained Optical Flow Estimation as a Matching Problem*, IEEE Trans. Image Proc. **22** (2013), no. 5, 2044–2055.
91. H.-H. Nagel, *constraints for the Estimation of Displacement Vector Fields from Image Sequences*, Proc. Int. Joint Conf. Artif. Intell., 1983, pp. 945–951.
92. H.-H. Nagel and W. Enkelmann, *An Investigation of Smoothness Constraints for the Estimation of Displacement Vector Fields from Image Sequences*, IEEE Trans. Patt. Anal. Mach. Intell. **8** (1986), no. 5, 565–593.

93. N. Parikh and S. Boyd, *Proximal Algorithms*, Found. Trends Optimization **1** (2013), no. 3, 1–108.
94. E. Parzen, *On the Estimation of a Probability Density Function and the Mode*, Ann. Math. Statistics **33** (1962), 1065–1076.
95. K. Prazdny, *Egomotion and Relative Depth Map from Optical Flow*, Biol. Cybernetics **36** (1980), 87–102.
96. N. Rao, B. Recht, and R.D. Nowak, *Signal Recovery in Unions of Subspaces with Applications to Compressive Imaging*, (2012), preprint arXiv:1209.3079.
97. R.T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, 2nd ed., Springer, 2009.
98. K. Rohr, *Landmark-Based Image Analysis*, Kluwer Acad. Publ., 2001.
99. H. Rue, S. Martino, and N. Chopin, *Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations*, J. Royal Stat. Soc. B **71** (2009), no. 2, 319–392.
100. P. Ruhnau, C. Gütter, T. Putze, and C. Schnörr, *A Variational Approach for Particle Tracking Velocimetry*, Meas. Science Techn. **16** (2005), 1449–1458.
101. P. Ruhnau, T. Kohlberger, H. Nobach, and C. Schnörr, *Variational Optical Flow Estimation for Particle Image Velocimetry*, Exp. Fluids **38** (2005), 21–32.
102. P. Ruhnau and C. Schnörr, *Optical Stokes Flow Estimation: An Imaging-Based Control Approach*, Exp. Fluids **42** (2007), 61–78.
103. P. Ruhnau, A. Stahl, and C. Schnörr, *Variational Estimation of Experimental Fluid Flows with Physics-Based Spatio-Temporal Regularization*, Meas. Science Techn. **18** (2007), 755–763.
104. Y. Saad, *Iterative Methods for Sparse Linear Systems*, SIAM, 2003.
105. S. Sabater, A. Almansa, and J. Morel, *Meaningful Matches in Stereovision*, IEEE Trans. Patt. Anal. Mach. Intell. **34** (2012), no. 5, 930–942.
106. B. Savchynskyy, J. Kappes, P. Swoboda, and C. Schnörr, *Global MAP-Optimality by Shrinking the Combinatorial Search Area with Convex Relaxation*, Proc. NIPS, 2013.
107. B. Schmitzer and C. Schnörr, *Modelling Convex Shape Priors and Matching based on the Gromov-Wasserstein Distance*, J. Math. Imag. Vision **46** (2013), no. 1, 143–159.
108. ———, *Object Segmentation by Shape Matching with Wasserstein Modes*, Proc. EMMCVPR, Springer, 2013.
109. C. Schnörr, *Determining Optical Flow for Irregular Domains by Minimizing Quadratic Functionals of a Certain Class*, Int. J. Comp. Vision **6** (1991), no. 1, 25–38.
110. ———, *Computation of Discontinuous Optical Flow by Domain Decomposition and Shape Optimization*, Int. J. Comp. Vision **8** (1992), no. 2, 153–165.
111. ———, *On Functionals with Greyvalue-Controlled Smoothness Terms for Determining Optical Flow*, IEEE Trans. Patt. Anal. Mach. Intell. **15** (1993), no. 10, 1074–1079.
112. N. Slesareva, A. Bruhn, and J. Weickert, *Optic Flow Goes Stereo: A Variational Method for Estimating Discontinuity Preserving Dense Disparity Maps*, Proc. 27th DAGM Symposium, 2005, pp. 33–40.
113. S. Sommer, F. Lauze, M. Nielsen, and X. Pennec, *Sparse Multi-Scale Diffeomorphic Registration: the Kernel Bundle Framework*, J. Math. Imag. Vision **46** (2013), no. 3, 292–308.
114. F. Steinbrücker, T. Pock, and D. Cremers, *Advanced Data Terms for Variational Optic Flow Estimation*, Proceedings Vision, Modeling and Visualization (VMV) (M. A. Magnor, B. Rosenhahn, and H. Theisel, eds.), DNB, 2009, pp. 155–164.
115. D. Sun, S. Roth, and M. Black, *A Quantitative Analysis of Current Practices in Optical Flow Estimation and the Principles Behind Them*, Int. J. Comp. Vision (2013).
116. O. Talagrand and P. Courtier, *Variational Assimilation of Meteorological Observations with the Adjoint Vorticity Equation. I: Theory*, Q. J. R. Meteorol. Soc. **113** (1987), no. 478, 1311–1328.
117. L. Tierney and J.B. Kadane, *Accurate Approximations for Posterior Moments and Marginal Densities*, J. Amer. Math. Soc. **81** (1986), no. 393, 82–86.
118. C. Tomasi and T. Kanade, *Shape and Motion from Image Streams under Orthography: a Factorization Method*, Int. J. Comp. Vision **9** (1992), no. 2, 137–154.
119. C. Trottenberg, U.and Oosterlee and A. Schüller, *Multigrid*, Academic Press, 2001.
120. P. Turaga, A. Veeraraghavan, A. Srivastava, and Chellappa. R., *Statistical Computations on Grassmann and Stiefel Manifolds for Image and Video-Based Recognition*, IEEE Trans. Patt. Anal. Mach. Intell. **33** (2011), no. 11, 2273–2286.
121. L. Valgaerts, A. Bruhn, M. Mainberger, and J. Weickert, *Dense versus Sparse Approaches for Estimating the Fundamental Matrix*, Int. J. Comp. Vision **96** (2012), no. 2, 212–234.
122. A. Verri and T. Poggio, *Motion Field and Optical Flow: Qualitative Properties*, IEEE Trans. Patt. Anal. Mach. Intell. **11** (1989), no. 5, 490–498.
123. C. Villani, *Optimal Transport: Old and New*, Springer, 2009.

124. P. Viola and W. M. Wells III, *Alignment by Maximization of Mutual Information*, Int. J. Comp. Vision **24** (1997), no. 2, 137–154.
125. A. Vlasenko and C. Schnörr, *Physically Consistent and Efficient Variational Denoising of Image Fluid Flow Estimates*, IEEE Trans. Image Proc. **19** (2010), no. 3, 586–595.
126. S. Volz, A. Bruhn, L. Valgaerts, and H. Zimmer, *Modeling Temporal Coherence for Optical Flow*, Proc. ICCV (D. N. Metaxas, L. Quan, A. Sanfeliu, and L. J. Van Gool, eds.), IEEE, 2011, pp. 1116–1123.
127. M.J. Wainwright and M.I. Jordan, *Graphical Models, Exponential Families, and Variational Inference*, Found. Trends Mach. Learning **1** (2008), no. 1-2, 1–305.
128. J.Y.A. Wang and E.H. Adelson, *Representing Moving Images with Layers*, IEEE Trans. Image Proc. **3** (1994), no. 5, 625–638.
129. J. Weickert and C. Schnörr, *A Theoretical Framework for Convex Regularizers in PDE–Based Computation of Image Motion*, Int. J. Computer Vision **45** (2001), no. 3, 245–264.
130. ———, *Variational Optic Flow Computation with a Spatio-Temporal Smoothness Constraint*, J. Math. Imaging and Vision **14** (2001), no. 3, 245–255.
131. P. Werkhoven and J.J. Koenderink, *Extraction of Motion Parallax Structure in the Visual System I*, Biol. Cybernetics **83** (1990), 185–191.
132. ———, *Extraction of Motion Parallax Structure in the Visual System II*, Biol. Cybernetics **63** (1990), 193–199.
133. P. Werkhoven, A. Toet, and J.J. Koenderink, *Displacement Estimates Through Adaptive Affinities*, IEEE Trans. Patt. Anal. Mach. Intell. **12** (1990), no. 7, 658–663.
134. M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof, *Anisotropic Huber-L1 Optical Flow*, Proc. BMVC (London, UK), 2009.
135. J. Westerweel, *Fundamentals of Digital Particle Image Velocimetry*, Meas. Sci. Techn. **8** (1998), 1379–1392.
136. G. Whaba, *Spline Models for Observational Data*, SIAM, Philadelphia, 1990.
137. B. Wirth, L. Bar, M. Rumpf, and G. Sapiro, *A Continuum Mechanical Approach to Geodesics in Shape Space*, Int. J. Comp. Vision **93** (2011), no. 3, 293–318.
138. K. Yamaguchi, D. McAllester, and R. Urtasun, *Robust Monocular Epipolar Flow Estimation*, Proc. CVPR, 2013.
139. L. Younes, *Shapes and Diffeomorphisms*, Appl. Math. Sciences, vol. 171, Springer, 2010.
140. L. Younes, F. Arrate, and M.I. Miller, *Evolution Equations in Computational Anatomy*, NeuoImage **45** (2009), no. 1, Suppl. 1, S40–S50.
141. J. Yuan, C. Schnörr, and E. Mémin, *Discrete Orthogonal Decomposition and Variational Fluid Flow Estimation*, J. Math. Imag. Vision **28** (2007), 67–80.
142. J. Yuan, C. Schnörr, and G. Steidl, *Simultaneous Optical Flow Estimation and Decomposition*, SIAM J. Scientific Computing **29** (2007), no. 6, 2283–2304.
143. ———, *Convex Hodge Decomposition and Regularization of Image Flows*, J. Math. Imag. Vision **33** (2009), no. 2, 169–177.
144. A.L. Yuille and N.M. Grzywacz, *A Mathematical Analysis of the Motion Coherence Theory*, Int. J. Comp. Vision **3** (1989), 155–175.

(F. Becker, S. Petra, C. Schnörr) IMAGE AND PATTERN ANALYSIS GROUP, UNIVERSITY OF HEIDELBERG, SPEYERER STR. 6, 69115 HEIDELBERG, GERMANY

E-mail address: {becker,petra,schnoerr}@math.uni-heidelberg.de

URL: ipa.iwr.uni-heidelberg.de