Computer Vision and Image Understanding 134 (2015) 1–21

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

# Optical flow modeling and computation: A survey

Denis Fortun *, Patrick Bouthemy, Charles Kervrann

*Inria, Centre de Rennes – Bretagne Atlantique, Rennes, France*

ARTICLE INFO

*Article history:*
Received 10 October 2014
Accepted 9 February 2015
Available online 26 February 2015

*Keywords:*
Optical flow
Motion estimation
Regularization
Parametric models
Optimization
Feature matching
Occlusions

ABSTRACT

Optical flow estimation is one of the oldest and still most active research domains in computer vision. In 35 years, many methodological concepts have been introduced and have progressively improved performances, while opening the way to new challenges. In the last decade, the growing interest in evaluation benchmarks has stimulated a great amount of work. In this paper, we propose a survey of optical flow estimation classifying the main principles elaborated during this evolution, with a particular concern given to recent developments. It is conceived as a tutorial organizing in a comprehensive framework current approaches and practices. We give insights on the motivations, interests and limitations of modeling and optimization techniques, and we highlight similarities between methods to allow for a clear understanding of their behavior.

© 2015 Elsevier Inc. All rights reserved.

## 1. Introduction

Motion analysis is one of the main tasks of computer vision. From an applicative viewpoint, the information brought by the dynamical behavior of observed objects or by the movement of the camera itself is a decisive element for the interpretation of observed phenomena. The motion characterizations can be extremely variable among the large number of application domains. Indeed, one can be interested in tracking objects, quantifying deformations, retrieving dominant motion, detecting abnormal behaviors, and so on. The most low-level characterization is the estimation of a dense motion field, corresponding to the displacement of each pixel, which is called optical flow. Most high-level motion analysis tasks employ optical flow as a fundamental basis upon which more semantic interpretation is built.

Optical flow estimation has given rise to a tremendous quantity of works for 35 years. If a certain continuity can be found since the seminal works of [120,170], a number of methodological innovations have progressively changed the field and improved performances. Evaluation benchmarks and applicative domains have followed this progress by proposing new challenges allowing methods to face more and more difficult situations in terms of motion discontinuities, large displacements, illumination changes or computational costs. Despite great advances, handling these issues in a unique method still remains an open problem. Comprehensive surveys of optical flow literature were carried out in the nineties [21,178,228]. More recently, reviewing works have focused on variational approaches [264], benchmark results [13], specific applications [115], or tutorials restricted to a certain subset of methods [177,260]. However, covering all the main estimation approaches and including recent developments in a comprehensive classification is still lacking in the optical flow field.

This survey paper is conceived as a tutorial aiming at organizing in a comprehensive framework the main approaches and practices for optical flow estimation. We will not try to tell the whole story of optical flow evolution, neither we give an exhaustive list of existing methods. We rather propose a synthetic classification of the main methodological principles of existing methods, with a particular concern given to recent developments. We will insist on the modeling aspects, practical interests and limitations of each introduced methodological element, with the aim of providing useful cues for a better understanding of optical flow. We adopt a classifying approach, decomposing the optical flow estimation problem in sub-parts. This viewpoint has the advantage of being didactic and facilitates the identification of similarities between methods in a global unique framework. However, this ambition of clarity should not hide that individual methods are often conceived as coherent approaches, and cannot be reduced to a mere assembly of isolated parts.

The remainder of this paper is organized as follows. In Section 2, we present general and basis concepts on optical flow used throughout the paper. Section 3 is devoted to a review of data constancy assumptions for optical flow and their usage for the design of data terms. In Section 4, the parametric approach, its specific issues and existing solutions are described. Section 5 presents

* Corresponding author at: Biomedical Imaging Group, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne VD, Switzerland.
*E-mail address:* denis.fortun@epfl.ch (D. Fortun).

http://dx.doi.org/10.1016/j.cviu.2015.02.008
1077-3142/© 2015 Elsevier Inc. All rights reserved.

the modeling and optimization aspects of globally regularized methods. Section 6 focuses on the occlusion issue. Section 7 is dedicated to the recent trend of integrating feature matching in the optical flow estimation process. Finally, the general conclusions of this survey are reported in Section 8.

## 2. Preliminaries

### 2.1. Definition of optical flow

The notion of *optical flow* literally refers to the displacements of intensity patterns. This definition originates from a physiological description of the visual perception of the world through image formation on the retina. In that sense, while optical flow is necessarily caused by relative motion between the observer and the objects of the observed scene, it only represents motion of intensities in the image plane, and not necessarily accounts for the actual 3D motion in the physical scene [251]. This definition of optical flow states that intensity of moving pixels remains constant during motion, and the main difficulty is to cope with the uncertainty due to the ill-posedness of the problem, called *aperture problem* (described in Section 2.4).

However, motion of interest in computer vision is in practice the real displacements of the objects. Thus, the variable that we actually want to retrieve is the projection on the image plane of the 3D motion in the scene, usually called *motion field*. The problem of recovering the 3D motion itself from several synchronized cameras is also an active research domain known as *scene flow* estimation [18,248]. The new problem arising is then that intensity changes are not necessarily caused by the displacements of objects of the scene, but can also be due to other disturbing phenomena like lighting changes, reflection effects or modifications of the properties of the objects affecting their light emission or reflectance.

The terms *optical flow* and *motion field* are usually mixed up to both designate the above definition of the *motion field*, and we will carry on this practice in the sequel of this survey.

### 2.2. Applicative domains

Optical flow provides fundamental information at the basis of various computer vision systems, in a wide range of applicative domains. We cite here some of the main fields where it is currently exploited.

The increasing use of multimedia devices involving video displays accentuates the importance of several video analysis problems. Action recognition is for instance an essential task for semantic interpretation of video content [1,130,256]. Video compression standards like MPEG exploit motion estimation to predict intermediate frames [131]. Optical flow can also be a determinant feature in video indexing and retrieval [123,202,230]. Video restoration of aged films is another field of interest, and can be improved by taking optical flow estimation into account [92,270].

In a biomedical context, dynamic properties of tissues or cellular objects are also of upmost importance. The deformation of organs [109,275] or the estimation of blood flow [124] are examples of medical applications that can require optical flow computation. Many conceptual and methodological links can also be found between medical image registration and optical flow [100,211,212,219,224]. In microscopy, dense motion can inform about cell deformation [3,90,139,203], motion of cellular structures [73,89], or help for individual cell tracking [167].

Robots or vehicle navigation also exploit optical flow as input of control systems for automatic guidance. Autonomous car driving has received particular attention in recent years [94,98,235]. Obstacle detection and avoidance are the main tasks investigated for general robot control in real environment exploiting optical flow [57,67,82].

Automated video surveillance is another growing research field requiring motion analysis. In particular, facial expression and gesture recognition [31,70], crowd motion and pedestrian behavior analysis [19,143] often involve dense motion analysis.

Among the other fields of interest, fluid flow analysis has applications in meteorology, oceanography aerodynamics or fluid mechanics and requires to consider specific physical constraints [65,113,115,168]. Several typical image deformations characterized by the conservation of texture properties during motion can be described as dynamic textures, and receive a specific treatment [69,85,84].

### 2.3. Evaluation of optical flow methods

As in other computer vision domains, much attention has been paid to the design of appropriate evaluation procedures for optical flow.

The visualization of motion fields provides a qualitative insight on the accuracy of the estimation. The two main visualization techniques are presented in Fig. 1. The arrow visualization directly represents motion vector and offers a good intuitive perception of physical motion. On the counterpart, a clean display requires to under-sample the motion field to prevent overlapping of arrows. The color code visualization associates a color hue to a direction and a saturation to the magnitude of the vector. It allows for a dense visualization of the flow field and a better visual perception of subtle differences between neighbor motion vectors.

On the other hand, objective quantitative evaluation based on error metrics is necessary for an accurate comparison of method performances. When ground truth is available, two error measures are commonly used, namely the Angular Error (AE) and the Endpoint Error (EPE). The AE of an estimated motion vector $w_{est} = (u_{est}, v_{est})^\top$ w.r.t. the reference vector $w_{ref} = (u_{ref}, v_{ref})^\top$ is defined by the 3D angle created by the extended vectors $(u_{est}, v_{est}, 1)^\top$ and $(u_{ref}, v_{ref}, 1)^\top$:

$$\mathrm{AE} = \cos^{-1}\left(\frac{u_{est}u_{ref} + v_{est}v_{ref} + 1}{\sqrt{u_{est}^2 + v_{est}^2 + 1}\sqrt{u_{ref}^2 + v_{ref}^2 + 1}}\right). \quad (1)$$

The EPE is defined as the Euclidean distance between the two vectors:

$$\mathrm{EPE} = \sqrt{(u_{est} - u_{ref})^2 + (v_{est} - v_{ref})^2}. \quad (2)$$

These two metrics are complementary regarding motion magnitude. AE is very sensitive to small estimation errors, occurring for small displacement, whereas EPE hardly discriminates between close motion vectors. On the contrary, AE tends to under-evaluate distances between motion vectors in some cases of large motion magnitudes, while EPE strongly penalizes large estimation errors.

The design of challenging benchmarks with ground truth on which to apply these error measures has motivated a substantial amount of work, which main steps are illustrated in Fig. 2. The first optical flow benchmark with ground truth established by [17] dealt either with simple parametric transformations applied to real images, like translation or rotation, or with synthetic sequences for which the true motion is available by construction. The resulting motion fields were characterized by small displacements and absence of discontinuities. More recent benchmarks have been designed to address new types of problems or specific applicative issues. The Middlebury benchmark[1] [11,13] is composed of more

[1] http://vision.middlebury.edu/flow/eval/.

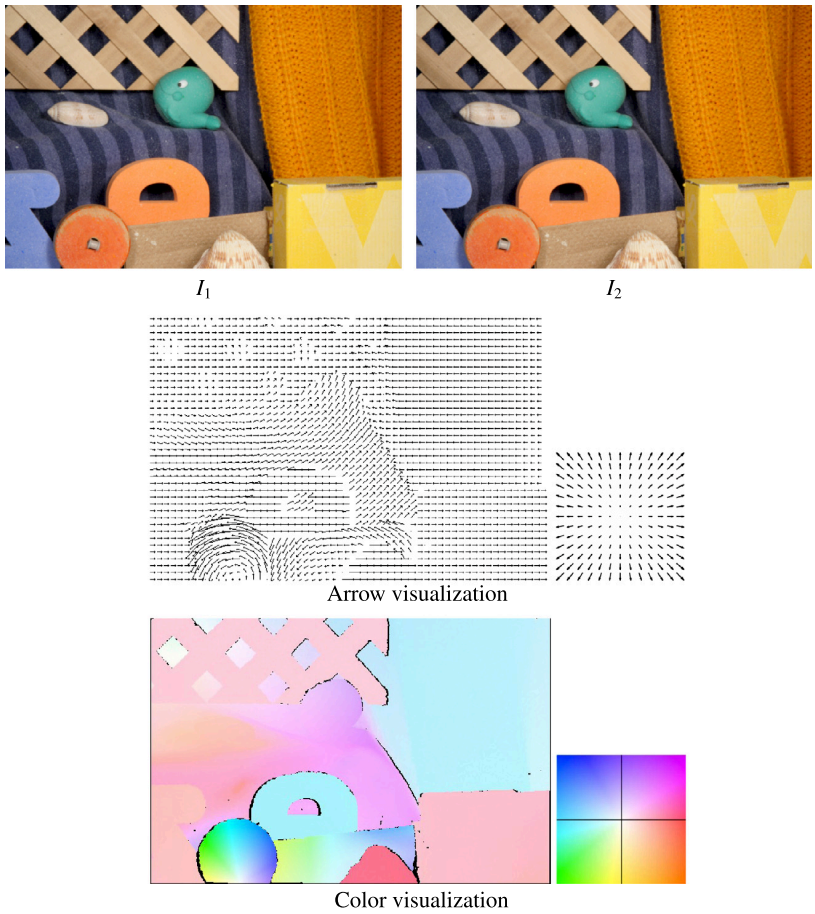


**Fig. 1.** Two types of visualization of the motion field transforming $I_1$ in $I_2$.

challenging sequences, partly made of smooth deformations similar to the sequences described in [17], but also involving motion discontinuities and motion details. While some sequences are synthetic, several others were acquired in a strictly controlled environment allowing to produce ground truth for real scenes. Issues raised by Middlebury being almost solved by modern methods, the MPI Sintel benchmark[2] [52] has been recently proposed. It is extracted from a synthetic movie and opens new issues mostly related to very large displacements, occlusions, illumination changes, and effects like blur or defocus. In parallel, dedicated datasets have been designed to solve specific problems related to applicative contexts, the most successful example being the KITTI benchmark[3] [94] devoted to assisted driving applications.

### 2.4. Estimation principles

Let us denote an image sequence by $I : \Omega \times T \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^2$ is the image domain and $T$ is the sampled time interval of the sequence. Every optical flow estimation method is based on an assumption on the relationship between the searched motion field $w : \Omega \to \mathbb{R}^2$ at time $t$ and the image $I(\cdot, t)$. The most natural and widely used assumption is that pixel intensity remains constant during displacement. The brightness constancy constraint equation (BCCE) is then defined by:

$$\frac{dI}{dt}(x(t), t) = 0. \tag{3}$$

Other feature constancy can be chosen, each encoding specific image properties, which will be discussed in Section 3. The discrete approximation of (3) at a given pixel $x \in \Omega$ and time $t$ yields:

$$I(x + w(x), t + 1) - I(x, t) = 0. \tag{4}$$

However, constraint (4) usually generates particularly difficult optimization problems. It can be much more tractable to consider the expanded version of (3) with partial derivatives, resulting in a linear version of (4):

$$\frac{\partial I}{\partial x_1}(x)\, u(x) + \frac{\partial I}{\partial x_2}(x)\, v(x) + \frac{\partial I}{\partial t}(x) = 0, \tag{5}$$

where $w(x) = (u(x), v(x))^\top$ and $x = (x_1, x_2)^\top$. For the sake of clarity, we have dropped the dependency of $I$ over $t$. We can also write (5) as

$$\nabla I(x) \cdot w(x) + I_t(x) = 0, \tag{6}$$

where $\nabla \cdot = \left(\frac{\partial \cdot}{\partial x_1}, \frac{\partial \cdot}{\partial x_2}\right)^T$ is the spatial gradient operator, $I_t = \partial I / \partial t$ is the partial temporal derivative and the dot "$\cdot$" denotes the inner product.

The linearized brightness constancy constraint (5) provides only one equation to recover the two unknown components of $w(x)$. From this single constraint, the component of the motion vector $w(x)$ in the direction of the image gradient can be computed, but

[2] http://sintel.is.tue.mpg.de/.
[3] http://www.cvlibs.net/datasets/kitti/index.php.

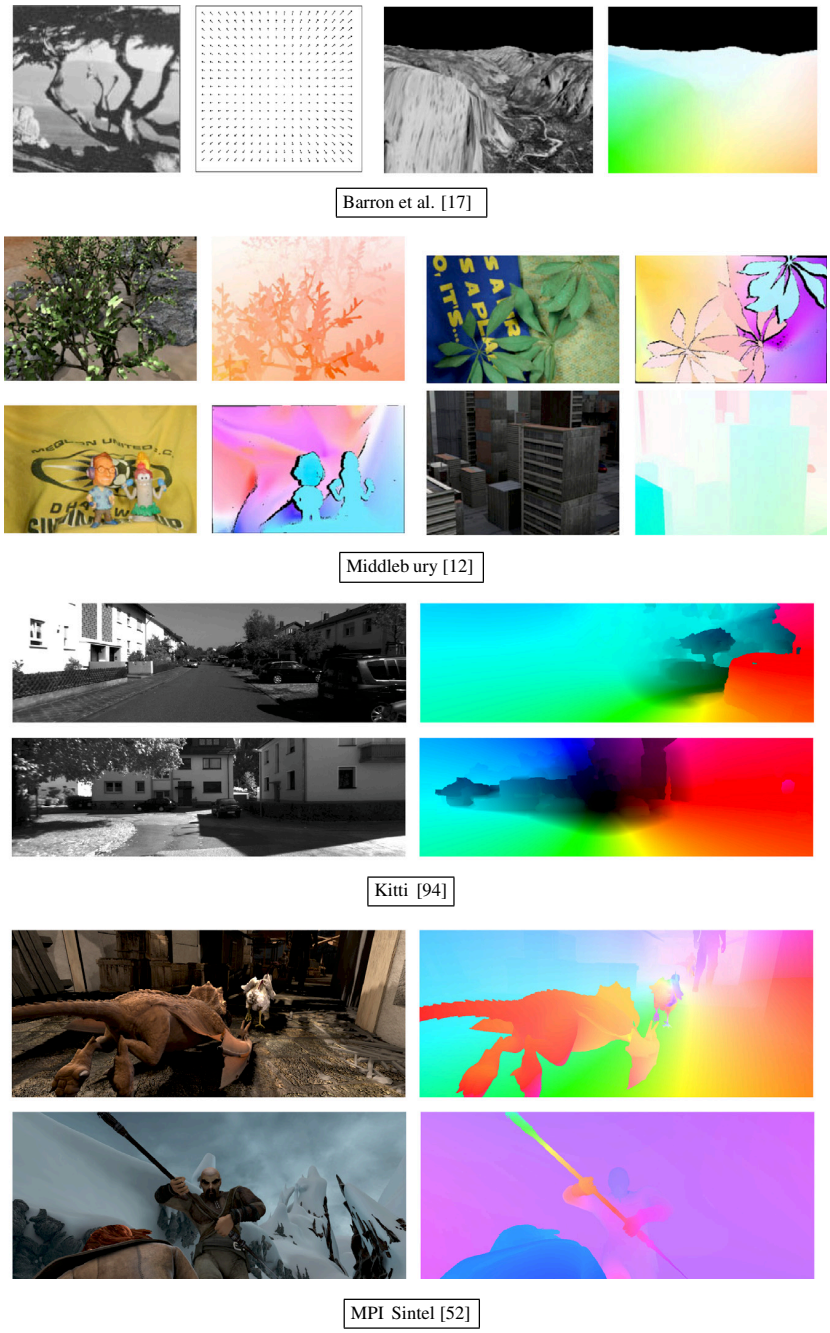


**Fig. 2.** Examples of frames and ground truth from the main optical flow evaluation benchmarks.

the two-dimensional problem remains under-constrained. This is known as the *aperture problem*, stating that motion of linear structures, as it is assumed by (5), is by nature ambiguous if the neighboring context is not taken into account.

To make the problem well-posed, it is necessary to introduce an additional constraint encoding *a priori* information on $w$. The *a priori* will take the form of spatial coherency imposed by either *local* or *global* constraints, described in Sections 4 and 5.

It is important to mention that while (4) holds for motion of arbitrary magnitude, the continuous motion constraint (5) restricts its validity domain to the linear region of $I$, which usually corresponds to small displacements or very smooth images. The linearization is nevertheless necessary for methods relying on differential computations. The standard technique to cope with large displacements is to embed the estimation in a coarse-to-fine scheme [27,81,174]. The idea is to create a pyramid of coarse-to-fine downsampled versions of the original image. At the coarsest level, the linearity domain of the image encompasses large displacements and the estimation can be based on (5). The estimations at coarser levels serves to warp the image at subsequent finer levels, where the estimation then reduces to a search for small motion increments. The solution is iteratively refined at each level until reaching the full image resolution. It is possible to continue the process in still higher resolutions created by interpolation, as proposed in the coarse-to-overfine approach [5]. The solution at each level of the multi-resolution pyramid can be interpreted as a fixed point in

the direct optimization of the non-linearized constancy Eq. (4) [45]. Almost all differential methods concerned with large displacements resort to the multiscale approach, possibly with some additional strategies to avoid its drawbacks.

The main undesirable effect produced by smoothing at coarse levels is the loss of small and rapidly moving objects in the final estimated flow field. If the object extent is smaller than its displacement, it is likely to be smoothed out at coarse levels and then "forgotten". Avoiding this drawback has been a very active topic in recent years. It has been accomplished mainly by integrating feature matching in the estimation process, as will be discussed in Section 7, or by resorting to discrete optimization methods, as detailed in Section 5.2.2.

## 3. Data cost

The deviations from the data constancy assumption (3) are penalized at every pixel $x \in \Omega$ via a data potential $\rho_{data}(x, I_1, I_2, w)$, defined in the case of brightness constancy (4) as

$$\rho_{data}(x, I_1, I_2, w) = \phi(I_2(x + w(x)) - I_1(x)), \tag{7}$$

where $I_1 = I(\cdot, t)$ and $I_2 = I(\cdot, t+1)$ denote two successive frames, and $\phi(\cdot)$ is the penalty function.

The brightness constancy assumption is in practice an imperfect photometric expression of the real physical motion in the scene. A common counter-example consists in moving the light source of an immobile scene, producing brightness variations without motion of any objects. In general, while it is possible to create synthetic sequences for which the constraint strictly holds [52], it is often violated in practice in case of changes in the illumination sources of the scene, shadows, noise in the acquisition process, specular reflections or large and complex deformation.

Choosing a quadratic penalty function $\phi(z) = z^2$, as in early works [120,170], makes optimization much easier, but assumes that the residual of the brightness constancy constraint Eq. (3) is normally distributed and thus gives a strong influence to large localized violations mentioned above. It is then common to resort to robust statistics [125] to reduce the impact of local errors considered as outliers. Common robust alternatives to quadratic penalty are the $L_1$ norm [45], the Tukey function [193], the Lorentzian norm [27] or the Leclerc's function [174]. Adapted optimization schemes must then be adopted to cope with non-linearity or non-convexity induced by the robust terms, as will be discussed in Sections 4.2 and 5.2. *A priori* smoothness assumption based on parametric constraint (Section 4) or explicit regularization (Section 5) also counterbalances local invalidity of data constancy.

Robust statistics and regularization treat the problem of violations of the constancy assumption by considering it as noise, with underlying distribution assumptions [155,221]. The considered distributions may not suitably model the possibly large localized violations implied by the above listed causes. Therefore, a large number of alternatives to brightness constancy have been proposed, aiming at more stable invariance properties. A few experimental studies have compared performances of different data costs given fixed optimization and regularization contexts [227,252].

Let us notice as a preamble to this section that it is difficult in practice to design a data term independently from the spatial coherence constraint and the optimization strategy to which it will be associated. For example, sophisticated feature conservation usually involves specific optimization difficulties, and thus requires careful design of the optimization solution. We will dedicate this section to a review of the main classes of data terms for optical flow estimation, emphasizing their validity domains and their limitations, independently from the estimation context in which they were elaborated.

### 3.1. Beyond brightness constancy

We explore several matching costs aiming at overcoming the drawback of the brightness constancy, in particular its sensitivity to noise and illumination changes.

#### 3.1.1. Image transformation

A first class of data potentials exploits the same pixel-wise form as (7), but operates on a transformed version $f(I)$ of the original image sequence:

$$\rho_{data}(x, I_1, I_2, w) = \phi(f(I_2)(x + w(x)) - f(I_1)(x)) \tag{8}$$

*Image smoothing*. We can first notice that Gaussian smoothing is applied as a pre-processing step by most methods [45,286], in order to reduce the influence of noise. It can be viewed as a modified version of brightness constancy, setting $f$ as a Gaussian filtering operator.

*High-order constancy*. Image derivatives possess illumination invariance properties that are well suited for motion estimation. The constancy of spatial image gradient, defined by $f(I) = \nabla I$, has been introduced in [245] for its ability to overcome the *aperture problem* when the determinant of the Hessian is non-zero. However, when applied on the directional derivative vectors, the gradient conservation only holds for translational or divergence motions. To achieve rotational invariance, the penalty should rather be applied on the magnitude of the derivatives, that is $f(I) = \|\nabla I\|$ [45]. It was subsequently used in the context of the *local* approach (see Section 4) in [240] and integrated in *global* variational methods (see Section 5) in [45].

Despite a demonstrated performance gain in the case of additive illumination changes compared to brightness constancy, gradient information is also much more sensitive to noise, and disappears in poorly textured regions. Therefore, it is always used in addition to brightness constancy. A large number of methods rely on this combination and achieve good results [45,181,279]. Finally [198] investigated higher-order data constancy like Laplacian of the image $f(I) = \Delta I$, or norm and determinant of the Hessian $f(I) = |\mathcal{H}I|, f(I) = \det(\mathcal{H}I)$.

*Texture*. Another way to obtain robustness against illumination changes is to work with the cartoon and texture components of the image, as initiated in [263]. The decomposition proposed in [8] consists in first obtaining the structure part $I_S$ by discontinuity-preserving smoothing (using the ROF model [210] in [263]), and then deriving texture part $I_T$ by subtracting $I_S$ to the original image. Additive illumination changes only affect the cartoon image while the texture image is less impacted. However, even if $I_T$ captures most image information, it is still insufficiently discriminative in some regions, and it is also more sensitive to noise than $I_S$. To limit this drawback, the texture image used to compute optical flow is usually blended with the cartoon part by a parameter $\gamma$ [154,231,261]: $f(I) = I_T - \gamma I_S$.

*Color spaces*. When dealing with color images, several photometric-invariant color spaces can be exploited. In particular, multiplicative illumination invariance is essential for realistic illumination models [246] and is achieved in the HSV space by the hue channel (local and global changes) and the saturation channel (only global changes) [176]. As for previously mentioned image transformations, the benefit in illumination change regions coincide with a loss of information in other parts, and the color channels are in practice combined with the intensity valued channel [286]. Other color spaces like normalized RGB [103] or spherical space [246] have been investigated.

*Combination of linear filters*. As mentioned before, it is often necessary to combine several constancy assumptions. It has been achieved in [259] by considering $K$ different linear filters $f_k$ providing a system of constancy equations to be solved at each pixel. In [232], a learning approach is proposed for finding the best combination of constraints. The data constraints induced by the linear filters $f_k$ are imposed through a Gaussian Scale Mixture (GSM) $\phi_{GSM}$.

$$\rho_{data}(x, I_1, I_2, w) = \sum_{k=1}^{K} \phi_{GSM}(f_k * I_2(x + w(x)) - f_k * I_1(x), \Xi_k), \quad (9)$$

where $*$ denotes the convolution operator and $\Xi_k$ are the parameters of the GSM associated to each filter. By learning the parameters of the GSM from a set of ground truth sequences, the "weights" of each filter in (9) are automatically determined. These weights are constant in the whole image. Spatially adaptive combination is also of upmost importance and will be addressed in Section 3.2.1.

#### 3.1.2. Patch-based similarity measures

Rather than pre-filtering images, neighborhood information can be integrated directly in the data term by patch-based similarity measures. Let us point out that a major issue with patch-based measure is the determination of the size an possibly shape of the patch.

*Filtering the data term*. In addition to pre-smoothing the images (8), Bruhn et al. [49] proposed to filter the data potential as follows:

$$\rho(x, I_1, I_2, w) = f(\phi(I_2(x + w(x)) - I_1(x))), \quad (10)$$

where $f$ is chosen as a Gaussian filter. While it was demonstrated beneficial for very noisy sequences, this choice also significantly blurs motion discontinuities and degrades the overall performance for low amount of noise, compared to pixel-wise data term, as emphasized in [286]. This limitation is addressed in [78,206] by replacing the Gaussian filtering with anisotropic discontinuity-preserving filtering (e.g. bilateral filtering in [78] and tensor voting in [206]). The relation between adaptive filtering and non-linear diffusion has also been exploited in [48].

*Correlation-based measures*. Similarity measures based on cross-correlation have been extensively used for various correspondence problems [44]. Normalized Cross Correlation (NCC) is usually preferred for its invariance to linear illumination changes. The NCC for a window $\mathcal{W}(x)$ centered at pixel $x$ is defined as

$$\mathrm{NCC}(x, I_1, I_2, w) = \frac{\sum_{y \in \mathcal{W}(x)} (I_2(y + w(x)) - \mu_2(x + w(x)))(I_1(y) - \mu_1(x))}{\upsilon_1(x)\upsilon_2(x + w(x))} \quad (11)$$

where for $i = \{1, 2\}$, $\mu_i(x)$ is the mean and $\upsilon_i(x)$ the standard deviation of $I_i$ in the window $\mathcal{W}(x)$. The associated data potential is

$$\rho_{data}(x, I_1, I_2, w) = 1 - \mathrm{NCC}(x, I_1, I_2, w). \quad (12)$$

NCC is actually discriminative enough to be used in a matching procedure without additional regularization, and produces coarse but reasonably robust motion fields. It is used in several applications like stereovision [72], fluid flow analysis [22] or biological imaging [146] where it also enables direct physical measures for diffusion processes.

The computational cost of (11) is a major limitation. Unlike simple cross-correlation which can be efficiently computed with Fast Fourier Transform (FFT), the computation of NCC for matching purpose cannot be easily performed in the frequency domain. In [163] only the numerator is computed with FFT and the denominator is rewritten as a product of sums, independent of the position of the pixel and thus efficiently computable with integral images [83]. In [171] this idea is generalized and the numerator is also computed with integral images, dramatically reducing the computation time and making it invariant to the patch size.

The challenge of integrating NCC in a variational optimization scheme is to cope with its non-differentiability. Indeed, Taylor expansion on the terms containing $w(x)$ in (11) still yields a highly non-linear potential. The approach of [180,272] has been applied to NCC but is able to handle arbitrary data terms as well. The authors directly linearize the data term and compute its spatial derivatives with finite differences. Werlberger et al. [272] keeps a second-order approximation to ensure the convexity of the energy, necessary in the primal–dual scheme used. Another recent technique allowing fast computation of NCC relies on the fact that NCC is actually equivalent to the Sum of Square Differences (SSD) when the images are filtered with the cheaply computed *correlation transform* [79].

*Census*. Census transform [283] recently regained interest and was promoted in [226] for optical flow estimation [108,118,179,182,183,205,252]. The Census signature is a bit string reflecting relative value of pixels of a patch with respect to the center pixel. By discarding the absolute intensity values, only the structure of the neighborhood is encoded in the signature, which makes it robust to illumination changes. It has shown robust behavior in outdoor scenes and vehicle driving scenarios [205,226,252]. Integrating the Census transform in variational optical flow is not trivial since it cannot be easily linearized. Solutions to remedy this problem are convex approximation [252], reformulation as a generalization of the gradient constancy conservation [108] or linearization of the data term [182,205] as previously mentioned for NCC [272].

### 3.2. Spatially adaptive data cost

The validity of each of matching costs is limited to a given range of visual situations. In a single frame, regions can coexist satisfying a given constancy assumption, and violating others. One solution could be to linearly combine them to take advantage of their complementary invariance properties. Softly selecting the best constancy constraint at each pixel is usually devoted to the robust penalty function, limiting the influence of local constraint violations. However, the data term should ideally be spatially adapted.

We distinguish two classes of methods achieving the spatial adaptivity: (i) optimization of the weights of a linear combination of data potentials, and (ii) estimation of the spatial distribution of the errors attached to a single data potential. The normalization of the data term used in [214,221,286] could fall in this category since a spatially varying weight is applied to the data term. It is derived from the linearized brightness constancy equation to prevent too strong constraints in regions of high image gradient (see a detailed interpretation in [286]).

#### 3.2.1. Combination of constancy assumptions

The combination of $P$ data constraints can be expressed as the weighted sum of their associated potentials $\rho_p(x, I_1, I_2, w)$:

$$\rho_{data}(x, I_1, I_2, w) = \sum_{p=1}^{P} \lambda_p(x)\, \rho_p(x, I_1, I_2, w). \quad (13)$$

Weights $\lambda_p(x)$ are spatially variant and have to be optimized to locally favor different data terms.

The idea of combining several data constraints has already been explored twenty years ago in [116]. In addition to the classical brightness constancy, the authors exploited a complementary sparse edge-based constraint. Weights $\lambda_p(x)$ are binary confidence measures derived from hypothesis testing providing evidence on each constraint.

In [279], intensity and gradient conservation are combined and their complementarity is experimentally demonstrated. The weights are defined to operate a binary selection between the two constraints and are obtained by considering a mean field approximation of (13), which intuitively amounts to selecting the constraint having the lower (normalized) potential. This idea has been used subsequently in [181].

The work of [140] addresses the problem in its most general form (13), allowing the combination of an arbitrary number and type of data conservation assumptions. A confidence measure for arbitrary data term is designed as an extension of the feature discriminability [218] to data discriminability. The confidence measures are used as local constraints on the weights $\lambda_p(x)$ in (13), and a regularization on $\lambda_p(x)$ is also imposed. The weights are then optimized jointly with the motion field.

#### 3.2.2. Modeling intensity variations

Another way to handle errors related to the constancy constraint is to design more elaborated models of intensity variations. The goal is to explicitly model the intensity changes due to environmental effects so as to isolate more accurately motion-induced changes.

The most common approach introduces an additional variable representing deviations from constancy assumption by a function $e(x, I_1, I_2, \xi)$ parametrized by the vector $\xi$:

$$\rho_{data}(x, I_1, I_2, w, \xi) = \phi(I_2(x + w(x)) - I_1(x) - e(x, I_1, I_2, \xi)). \tag{14}$$

The model proposed by [188] is composed of an offset change $\xi_o : \Omega \to \mathbb{R}$, accounting, e.g., for moving shadows or highlights, and a multiplicative change $\xi_m : \Omega \to \mathbb{R}$ encoding linear illumination variations. The error function can then be expressed as

$$e(x, I_1, I_2, \xi) = \xi_m(x) I_1(x) + \xi_o(x). \tag{15}$$

This general formulation has been exploited in a number of works, considering either the offset parameter alone [55,89,193], the multiplier alone [284] or both parameters [142,159,239]. They may differ on the type of spatial coherency, the penalty function, or the optimization strategy. A smoothness constraint is assumed on $\xi_o$ and $\xi_m$, either with a local parametric form [188,193] or a global regularization [55,89,142,159,239]. The offset formulation is also used in [10,88], but is associated to a sparsity constraint aiming at retrieving violations due to occlusions. The more general approach of [74] parametrizes intensity changes in terms of Brightness Transfer Function [106], which coefficients are learned from training data.

The model (15) is based on a general polynomial approximation. If specific knowledge about the observed physical process is available, dedicated models can be designed. A number of physical constraints are explored in [110], where a generic local estimation framework is designed based on a Taylor expansion of arbitrary data constraints similar to the subsequent methods [180,272].

## 4. Parametric approach

As explained in Section 2.4, a spatial coherence constraint on the flow field must be added to the previously described data terms. To this end, one can impose the flow field to follow a parametric model in region $\mathcal{R} \subseteq \Omega$. The motion field $w_\alpha : \mathcal{R} \to \mathbb{R}^2$ is then fully characterized by the associated parameter vector $\alpha$. When the region $\mathcal{R}$ is a small sub-domain of the image, these methods are referred to as *local* approaches. The objective energy to be minimized is the weighted sum of the potentials provided by each pixel of $\mathcal{R}$:

$$\widehat{\alpha} = \arg\min_{\alpha} \sum_{x \in \mathcal{R}} g(x)\ \rho_{data}(x, I_1, I_2, w_\alpha) \tag{16}$$

where $g(x)$ is a spatial weighting function controlling the influence of pixel $x$ in the estimation.

It is crucial to determine the local estimation domain where the parametric form of the motion model is a valid approximation of the true motion. Low-order polynomial motion models like translation or affine deformation can usually represent motion in small neighborhoods, whereas more complex models like deviations from affine constraint or combination of basis functions can deal with larger regions.

We first give an overview of the mostly used motion models and their associated optimization strategies. Secondly, we discuss about the different ways to define appropriate local estimation supports.

### 4.1. Motion models

The choice of the motion model is driven by a trade-off between efficiency and representativeness. Complex nonlinear and physical-based models can be exploited to model deformations for image registration [224]. These models are particularly well adapted to physically constrained situations as they can be encountered in medical imaging, and in particular to capture smooth deformations. In contrast, optical flow is dealing with temporal sequences of arbitrary content, usually involving motions of several objects with unrelated behaviors, generating motion discontinuities as well as smooth motion parts. As a result, it is difficult to capture the whole complexity of motion fields with a single unifying and computationally tractable parametric model. Therefore, attempts in this direction are not frequent and not among the best performing methods in optical flow benchmarks. The approach of most parametric methods for optical flow is rather to rely on much simpler motion models, mostly polynomial models, and to restrict their application to local domains, where they can represent accurate approximations.

We will restrict ourselves to linear models of the form

$$w_\alpha(x) = \sum_{k=1}^{K} b_k(x) \alpha_k(x), \tag{17}$$

where $b = \{b_k(x)\}_{k \in [1,K]}, b_k(x) \in \mathbb{R}^2$ are basis functions and $\alpha(x) = \{\alpha_k(x)\}_{k \in [1,K]}$ are the parameters to be optimized. Other parametric models than those described here can be found, like planar surfaces or rigid body [23], or wavelet bases [75,76,217,274]. It can also be noted that parametric models are sometimes completed with explicit regularization terms (see Section 5) imposed on the parameters themselves [75,133,175,191].

#### 4.1.1. Polynomial models

Polynomial models are among the most compact parametric representations of motion fields and are also remarkably well suited to retrieve local physical motion of individual objects and even deformable motion. The polynomial model can be seen as a special case of (17), which we write for clarity as:

$$w_\theta(x) = \mathbf{B}(x)\ \theta, \tag{18}$$

where $\mathbf{B}(x)$ is a matrix determining the form of the model and $\theta$ is the parameter vector. Apart from the exception of [191] where the parameters are spatially variant and regularized, the parameter $\theta$ is kept constant over the estimation domain. The matrix $\mathbf{B}(x)$ depends on the pixel coordinates $x \in \mathcal{R}$, and the order of the polynomial determines the complexity of the motion field. Low-order polynomials are usually sufficient to model smooth motion fields, and their small number of parameters allows for efficient computation. The two mostly used polynomial models correspond to the translation and the affine motion:

$$\text{Translational}: \theta = (a_1, a_2)^\top; \quad \mathbf{B}(x) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\text{Affine}: \theta = (a_1, a_2, a_3, a_4, a_5, a_6)^\top; \quad \mathbf{B}(x) = \begin{pmatrix} 1 & x_1 & x_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_1 & x_2 \end{pmatrix}$$

with $x = (x_1, x_2)^\top$.

The translation assumption is very restrictive and must be applied to very small regions [170]. The physical assumption underlying the affine model is a rigid motion of 3D planar surfaces projected orthogonally on the image plane. Higher-order polynomials can model more complex situations, but are still too smooth to allow for motion discontinuities. For example, the 8-parameter quadratic model represents rigid motion of a planar surface in perspective projection. The small number of parameters of the affine model and its realistic local assumption make it often considered as the best trade-off between complexity and descriptiveness [27,88,175,193]. The accuracy of local affine motion estimation in appropriate estimation support was experimentally demonstrated in [88].

#### 4.1.2. Learned basis

The basis functions can also be learned from a set of training flow fields. As for polynomial models, the resulting learned motion models cannot describe motion of complex scenes, but they are able to retrieve a larger diversity of local motion patterns, including discontinuities.

The design of the training set reflects the assumption on the form of the flow. In a generic point of view, Black et al. [32] used synthetic motion fields representing simple motion patterns. The method of [190] relies on a large number of patches of ground truth motion fields. The training set can be dedicated to a specific application, as in [87] where the aim is to estimate mouth motion. To avoid resorting to external ground truth, [93] defines the training set on the processed sequence itself. The basis set is composed of trajectories constructed by feature tracking on large temporal scales, in regions ensuring reliable tracks. In all these works, an orthogonal basis of flow fields is generated by PCA decomposition, conserving only the first $K$ components containing most of the variance of the training set.

#### 4.1.3. Free-form deformations

The free-form deformation model (FFD) [211] has been originally introduced for image registration and has demonstrated great efficiency to retrieve smooth deformations. The displacements are defined on a coarse regular subgrid of the image and is interpolated on the final resolution with B-splines. The motion basis $b_k(x)$ is thus formed with the displacements of the $K$ control points and coefficients $\alpha_k(x)$ are B-spline influence functions. The dimensionality reduction induced by the subsampling of the image grid makes the computation much easier, and the spatial coherence of the deformation is ensured by the B-spline interpolation. On the counterpart, the framework cannot retrieve sharp motion discontinuities, while it is necessary for optical flow applications. Image-adaptive non-regular control points distribution [212] or coarse-to-fine spacing strategies [211] are possibilities to address this issue.

The work of [236] was the first to apply this idea to optical flow, with non-uniform control points defined on image-driven quadtrees. In [201], B-Splines defined on a uniform grid are used to retrieve smooth deformations. In [101] the problem is turned in a discrete setting. The range of motion labels is iteratively adapted by estimating a local uncertainty covariance. Despite good results, the method is still limited by over-smoothing. The method of [219] addresses the discontinuity problem with a sparsity constraint on the B-spline coefficients, allowing to modulate the influence of the control points. Recent approaches [100,101,219] also add an explicit regularization on the motion field, overlapping with the methodology described in Section 5.

### 4.2. Optimization

Parametric models are usually associated with the penalty of a pixel-wise data constancy constraint (7). In case of intensity constancy (5), energy (16) then writes:

$$E(\alpha) = \sum_{x \in \mathcal{R}} g(x) \phi(\nabla I(x) \cdot w_\alpha(x) + I_t(x)) \tag{19}$$

where $\nabla I$ is the image gradient and $I_t = \partial I / \partial t$. The special case of a quadratic function $\phi$ and a translational model as in [170] leads to a very simple optimization problem, since the cancelling of the derivatives of (19) amounts to solving the linear system $\mathbf{M}\alpha = b$, with

$$\mathbf{M} = \begin{pmatrix} \sum_x g(x) I_{x_1}^2(x) & \sum_x g(x) I_{x_1}(x) I_{x_2}(x) \\ \sum_x g(x) I_{x_1}(x) I_{x_2}(x) & \sum_x g(x) I_{x_2}^2(x) \end{pmatrix}$$

$$b = -\begin{pmatrix} \sum_x g(x) I_{x_1}(x) I_t(x) \\ \sum_x g(x) I_{x_2}(x) I_t(x) \end{pmatrix}$$

where $I_{x_1}$ and $I_{x_2}$ are the partial derivatives of $I$ respectively along the horizontal axis $x_1$ and the vertical axis $x_2$. The rank of $\mathbf{M}$ allows one to decide if a unique solution of the linear system exists, and can be used to adapt the size of the local domain $\mathcal{R}$ (see Section 4.3.2). Despite the limitations of the quadratic penalty, this approach has become very popular for its implementation simplicity, low computational cost and available code in the OpenCV library [36,41].

However, robust estimation is often advocated [27,32,76,193,215] as mentioned in Section 3, especially for polynomial models, to deal with the frequent case of multiple motions in the estimation domain. Among the variety of optimization methods used in case of robust penalty function, the Iterative Reweighted Least Squares (IRLS) [119] and gradient descent approaches have mostly been used. IRLS proceeds by successive optimizations of quadratic problems weighted by a function of the current estimate and is implemented in the Motion2D software [193,215] with. Gradient descent approaches are often coupled with Graduated Non-Convexity (GNC) [27,33] to cope with non-convexity of (19). Regarding the slow convergence of steepest descent, it is preferable to use second order approximations and Newton methods, or quasi Newton methods like L-BFGS or Levenberg–Marquardt, approximating the Hessian for large dimension problems.

### 4.3. Neighborhood selection

As previously mentioned, the spatial adaptation of the parameters $\alpha(x)$ is a way to cope with complex and discontinuous flow fields [93,191,219]. This approach often involves *a priori* constraints on the spatial distribution of $\alpha(x)$ and is thus strongly related to the methods that will be presented in Section 5. Such a dense parameter map moves away from the compactness of parametric models.

On the other hand, when parameters are constant over the estimation domain $\mathcal{R}$, the resulting motion field is smooth and can constitute a valid approximation only in regions of coherent motion. The region $\mathcal{R}$ must be large enough to enable motion estimation, while small enough to keep valid the parametric approximation (*generalized aperture problem* [27]). We will

describe below the strategies for defining $\mathcal{R}$ in the case of constant parameters $\alpha(x) = \alpha$ over $\mathcal{R}$, for polynomial models.

#### 4.3.1. Entire domain

Despite their inability to retrieve realistic motion in arbitrary scenes, polynomial models combined with robust estimation are well adapted to capture the dominant image motion. Applied in the whole image domain $\Omega$, they become particularly useful to separate or compensate the camera motion [193]. Applications to moving object detection [194] or to action recognition [202] are among the most common ones.

#### 4.3.2. Square patches

The approach initiated by [170] performs independent estimations in small square or circular patches. Most of the related methods use fixed patch size, and conserve the velocity vector deduced from the estimated motion model at the square center [12,26,141,284]. This choice is very popular for its simplicity of implementation and can be naturally parallelized [222,284]. It is still extensively used for numerous applications. However, following the generalized aperture problem, patches centered at each pixel with a fixed size are likely to contain either multiple motions or no image gradient.

Multiple motions in a single patch can be partially handled with robust estimation by rejecting secondary motions, considered as outliers [27,95,193,215].

The second option is to adapt the size or the position of the patch so that it contains an unimodal motion distribution. The size of the patch can be adapted with a bias-variance criterion [173], or based on a confidence measure on the reliability of the local domain for parametric estimation. Starting from a small patch size, it is thus possible to increase the size of the patch until the condition for a reliable domain is violated [173,215]. In the translational model case, one can analyze the singularity of the matrix $\mathbf{M}$ of the linear system (20) and, for example, impose a minimum threshold to the maximum eigenvalue of $\mathbf{M}$ [17]. Rather than adapting the size, [132] adapts the position of the patches. Patches corrupted by strong intensity edges are displaced by a mean-shift procedure to reach more homogeneous regions. The method described in [88] adapts both size and position of the square patches. This adaptation is performed thanks to the aggregation framework decoupling the estimation of candidates with predefined sizes and positions, and their selection achieved with a global model (see Section 5).

Finally, local adaptation of the weights $g(x)$ in (16) is another way to adapt the estimation support. It can be done for instance with bilateral filter weights as in [78], or using the non-linear structure tensor [48].

#### 4.3.3. Segmented regions

The optimal supports to perform polynomial motion estimation ideally correspond to a segmentation of the image in coherently moving regions. We briefly describe two types of approaches: independent image segmentation and joint estimation of motion and region supports or frontiers.

*Image segmentation*. While the ultimate goal is to segment the unknown motion field, color-based image segmentation is a much simpler alternative which can help motion estimation. It can be reasonably assumed that motion discontinuities coincide with image discontinuities (but the inverse is far from being true). It implies that an image segmentation is a motion field over-segmentation, and obtained regions are thus guaranteed to contain no motion discontinuity. However, merely estimating motion in the resulting regions is problematic for two reasons.

The first limitation is that the segmented regions may not contain enough information for motion estimation. Parametric estimations in these regions must be performed by circumvented ways. The very fine over-segmentation of [287] imposes for instance to perform region matching. Generally, an independent coarse and cheap motion estimation is fused with the color image segmentation to overcome the lack of information [28,34,278]. In [278], hybrid regions are found by applying mean-shift segmentation in the extended space of color and motion. Differently, [28,34] fit a parametric flow field on the coarse initial motion field, obtained with a global regularized method [30] for [28], and with the sparse KLT tracker [218] for [34].

The second problem is that spatial coherence between estimated motion in neighboring segments is not ensured. Global regularization (see Section 5) can here be imposed, either on the motion parameters associated to each region [278] (similarly to [133], not resorting to image segmentation), on the coarse motion field completing color information [28] or, in a layered approach, on the layer assignment function [34].

*Joint estimation and segmentation*. Color segmentation is usually too dependent on the image content to make it the basis of a robust motion estimation method. Rather than considering segmentation and estimation as two independent tasks, most methods have a coupled approach of the problem. Motion parameters and region supports are jointly estimated by minimizing a global energy imposing a coupling between them. This approach has first been addressed as a labeling problem [37,195] where the label field $l : \Omega \to \{l_1, \dots, l_N\}$ associated to the $N$ regions is estimated jointly with the motion parameters in each region $\alpha = \{\alpha_1, \dots, \alpha_N\}$, in a discrete Markov Random Field framework:

$$E(\alpha, l) = \sum_{x \in \Omega_d} \rho_{data}\left(x, I_1, I_2, w_{\alpha_{l(x)}}\right) + \sum_{\langle x,y \rangle} \rho_{reg}^{\mathrm{MRF}}(l(x), l(y)), \tag{20}$$

where $\Omega_d$ is the discrete image domain and $\rho_{reg}^{\mathrm{MRF}}(l(x), l(y))$ is a regularization prior on the label field, typically chosen as $\rho_{reg}^{\mathrm{MRF}}(l(x), l(y)) = 1 - \delta(l(x), l(y))$, with $\delta(\cdot)$ the Kronecker function. The optimization is done alternatively between motion and regions. Another viewpoint in a variational framework extends the Mumford–Shah formulation of image segmentation [184] to motion segmentation [66,199,247]. In addition to the data fitting potential inside each region $\mathcal{R}_i$, a constraint restricting the length $\mathcal{L}(C)$ of the set of region boundaries $C$ is imposed globally, resulting in the energy

$$E(\alpha, C) = \sum_{i=1}^{N} \int_{\mathcal{R}_i} \rho_{data}(x, I_1, I_2, w_{\alpha_i}) dx + \nu \mathcal{L}(C). \tag{21}$$

The minimization is performed alternatively on the flow and the boundaries. Minimizing (21) with respect to $C$ requires a differentiable approximation of the contour length $\mathcal{L}(C)$. It is common to implicitly represent the partitioning of the image with level sets, which allows to represent the interior of the regions by the sign of the function, as well as the total length of the boundaries by their level lines. One level set function can only represent two regions. For an arbitrary number of $N$ regions, it is possible to define $N$ corresponding levels sets, at the price of a high computational cost and a more complex energy to prevent vacuums in the partitioning. Other strategies can be employed, as the one of [56] re-used in [66], for more sophisticated combinations between functions. In [199], this level set framework is augmented with an edge-driven tracking and background detection. In [207], the two images are jointly segmented, and they influence each other through a dynamic prior term determined by the similarity of the two evolving shapes. Segmentation of static and dynamic textures is performed in [84] by assigning different data conservation assumptions to each class. A graph-cut optimization scheme has been proposed in [80]. These two formulations (20) and (21) can be found in numerous other works [137,175,213,244]. Similarly, layered approaches

[9,223,231,234,257,277] aim at decomposing the observed scene into overlapping layers and introduce information about depth ordering of regions. An important interest of such methods is to provide a natural occlusion handling framework.

Two main drawbacks affect the joint motion estimation and segmentation approach. First, alternating minimization w.r.t. regions and flow fields is computationally expensive. The related layered approach [233,234] achieving state-of-the-art results requires several hours to process a pair of $640 \times 480$ pixels, and even GPU-based implementation [244] can need up to an hour.

Second, the initialization of motion and segmentation parameters usually has a substantial impact on the evolution of the region contours throughout the minimization procedure. Therefore, paradoxically, the best performing methods in terms of accuracy of optical flow fields resort to pre-calculated motion fields obtained with independent methods to initialize their algorithm. For example, the results reported in [233,244] are obtained with an initialization by the result of the methods of [231,271] respectively. Nevertheless it is shown in [244] that the initialization affects more the speed of convergence than the final result.

Some works exploit more complex flow field representations than polynomial approximation, by authorizing deformations from an initial affine model [28,175,233] or an explicit regularized model [4,46]. Allowing such complex and discontinuous motion fields in segmented regions actually tends too produce larger regions, and ultimately leads to *global* approaches where motion discontinuities are handled by the global modeling itself. Consequently, regions may not represent coherent motion, but rather a delineation adapted to the specific estimation method. Finally, imposing prior on shapes and contours has been investigated in medical imaging [138], but a limited number of regions can be handled at the same time.

## 5. Regularized models

In an alternating manner to the parametric representation of spatial coherence, mostly adapted to smooth deformations, the global form of the motion field can be imposed by an explicit regularization term. Motion discontinuities are then no more represented by the boundaries of the regions delimiting parametric motion fields, but they are involved in the global model, often considered as outliers w.r.t. smoothness assumptions. The variational approach has been initially proposed by [120] and is usually referred to as the *global* approach, since the regularization term interconnects all the pixels of the image and thus requires the optimization of the objective energy to be performed globally. In this subsection, we review current versions of the regularization model and optimization strategies.

In its most general form, the energy minimized by global regularization methods can be written as:

$$E_{\text{global}}(w) = \int_{\Omega} \rho_{data}(x, I_1, I_2, w) + \lambda \rho_{reg}(x, w) dx \tag{22}$$

where $\rho_{data}(x, I_1, I_2, w)$ is the data potential, as discussed in Section 3, $\rho_{reg}(x, w)$ is the regularization potential encoding an *a priori* assumption on the field $w$, and $\lambda$ is a parameter tuning the balance between the two terms. Broadly speaking, the regularization potential aims at smoothing the motion field in regions of coherent motion while preserving motion discontinuities at the boundaries of moving objects. Finding the trade-off can also be partially addressed in the adaptation of the balance parameter $\lambda$ [112,156,189,286].

A major interest of the global variational framework is its versatility, allowing one to model different forms of flow fields by combining different data and regularization terms. One must nevertheless keep in mind that minimizing (22) is often a tricky task. The potentially unlimited combinations of data terms and regularization terms is restricted in practice to those compatible with efficient minimization. Besides, advances in optical flow have often been correlated with new possibilities offered by optimization techniques. For example, efficient Primal–Dual minimization for Total Variation regularization [55] have motivated a number of optical flow models [244,269,285]. The development of efficient discrete optimization techniques based on graph cuts [40] or message passing [147] also inspired various works [59,161,181]. Another consequence of the close intricacy between energy model and optimization technique is the difficulty to compare performances of different models, as global optimum is in general not guaranteed for sophisticated energies and the quality of the local optimum depends on the type of optimization method.

We will detail in Section 5.1 existing regularization models independently from optimization techniques, for the sake of clarity. Section 5.2 focuses on the dependency between specific energy models and optimization methods.

### 5.1. Regularization models

#### 5.1.1. Spatial flow gradient constraint

The most natural and widely used way to impose smoothness of the motion field is to penalize the magnitude of the flow gradient:

$$\rho_{reg}(x, w) = h(x, I_1)\, \phi(\|\nabla w(x)\|^2) \tag{23}$$

where $\phi(\cdot)$ is the penalty function and $h(x, I_1)$ is a weighting function.

A taxonomy of optical flow regularizers has been proposed in [266]. The authors focus on convex and rotational invariance properties, and prove uniqueness of the solution in each case. For each regularization of type (23), they show the equivalence between the resolution of the Euler–Lagrange equations associated with energy (22) and diffusion filtering. In addition, a diffusion tensor is derived for each particular variation of (23). We will give a more succinct overview, taking only some elements from this classification and integrating more recent approaches.

*Flow-driven regularization.* In flow-driven approaches, no relation between the form of the flow field and the structure of the image is assumed. The weighting function is thus $h(x, I_1) = 1,\ \forall x \in \Omega$. The seminal formulation of [120] adopts a quadratic penalization function:

$$\rho_{reg}(x, w) = \nabla u(x)^2 + \nabla v(x)^2, \tag{24}$$

with $w(x) = (u(x), v(x))^\top$. The quadratic penalization being unable to capture motion discontinuities, the introduction of a complementary line process to handle motion discontinuities in the MRF framework [116,153], and the use of robust sub-quadratic penalties [27,77,174] have soon been employed to overcome the problem. Among the wide panel of robust functions, the popular parameter-free Total Variation (TV) prior, has interesting and useful properties [45,63,279,285]. Contrary to most other robust norms, the TV yields a convex constraint facilitating optimization. The non-differentiability in 0 is generally alleviated by using the regularized version $\phi(z) = \sqrt{z^2 + \epsilon^2}$, where $\epsilon$ is a small constant. Associated to proximal splitting minimization, TV involves solving several ROF (Rudin–Osher–Fatemi) models [210], for which very efficient algorithms exist [54]. A series of optical flow estimation methods have exploited this idea for fast and accurate minimization [55,263,272,285].

TV regularization actually favors piecewise constant flow fields. This framework is known to transform a smoothly varying variable to a succession of small discontinuous constant steps (*staircasing* artifacts). This undesirable effect can be reduced by replacing the

$L_1$ penalization by a quadratic one for small gradient magnitude, which is the behavior of the Huber norm [220,271]. Another possibility is to penalize higher-order derivatives of the flow, as done in [241] for the second derivative, to favor piecewise affine flow fields. The Total Generalized Variation (TGV) [43], generalizes $L_1$ penalization to arbitrarily high order derivatives. The performance gain of the second order TGV has been experimentally shown for smooth deformation conditions, for which the *staircasing* effect is prominent with TV regularization [42,205,252].

Despite the demonstrated performance of TV due to its algorithmic attractiveness, the real distribution of optical flow derivatives has been shown to follow a more heavy-tailed and concave distribution [208]. Finding good approximate solutions for non-convex priors has motivated a number of works and will be discussed in Section 5.2. When appropriate minimization strategy is available, like graph cuts or the recent work of [192], this kind of penalty functions has proven to yield improvements compared to the TV model. Efficient methods to minimize Potts model have also recently been investigated [53].

*Inclusion of image gradient information.* It is natural to assume a link between the motion field and its source image $I_1$. As already stated in Section 4.3.3 about the relationship between motion and image segmentation, it is reasonable to consider that motion discontinuities coincide with image discontinuities delineating moving objects. This information can be incorporated in the regularization through the weighting function $h(x, I_1)$ taken as a smooth decreasing function of $\|\nabla I\|^2$ [2,10,181,263,279], often defined as

$$h(x, I_1) = e^{-\|\nabla I(x)\|^2/\varsigma^2}, \tag{25}$$

where $\varsigma$ is a parameter setting the influence of the image gradient on the regularization. Despite the risk of over-segmentation, this simple weighting strategy usually improves experimental results.

The weighting function (25) is isotropic since the smoothing is modulated by the same value in all the directions. This is suboptimal since we would ideally like to prevent the smoothing only *across* the boundaries, and allow it *along* them. This can be achieved by defining the regularization axes differently from the horizontal and vertical axes. The eigenvectors $s_1$ and $s_2$ of the structure tensor $\mathbf{R}_\rho(x) = \mathbf{K}_\rho * [\nabla I(x)\nabla I(x)^\top]$ are well adapted since $s_1$ is oriented across local image edges and $s_2$ is orthogonal to $s_1$. This idea has been first introduced in [186], which regularizer has been rewritten in [286] as:

$$\rho_{reg}(x, w) = \frac{1}{\|\nabla I(x)\|^2 + 2\kappa^2}\Big(\kappa^2\Big(u_{s_1}^2(x) + v_{s_1}^2(x)\Big) + (\|\nabla I(x)\|^2 + \kappa^2)\Big(u_{s_2}^2(x) + v_{s_2}^2(x)\Big)\Big) \tag{26}$$

where $u_{s_i}, \{i = 1, 2\}$, are the derivatives of $u$ along the $s_i$ axis and $\kappa$ is a regularization parameter. When $\kappa$ is small, the regularization is reduced in the direction of the image gradient $s_1$ and strengthened along image edges $s_2$ depending on the image gradient magnitude $\|\nabla I(x)\|^2$. In [186], the eigenvectors $s_1$ and $s_2$ are obtained with a radius $\rho = 0$ for the Gaussian filtering of the structure tensor $\mathbf{R}_\rho(x)$.

The classical artifact produced by purely image-driven regularization is an over-fitting of the flow field on image boundaries, creating artificial motion discontinuities. To reduce the impact of image gradient in the regularization, Sun et al. proposed in [232] to keep the $s_1$ and $s_2$ directions, while suppressing the weighting on $\|\nabla I(x)\|^2$ in (26) and employing a robust penalty function $\phi$:

$$\rho_{reg}(x, w) = \phi(u_{s_1}^2(x)) + \phi(v_{s_1}^2(x)) + \phi(u_{s_2}^2(x)) + \phi(v_{s_2}^2(x)). \tag{27}$$

The work of [286] proposed a generalized computation of the regularization axes, oriented to follow the data constraint rather than the image edges. In analogy with the previous approach defining $s_1, s_2$ from the structure tensor, they compute the eigenvectors of a so-called *regularization tensor*, designed to be complementary with the data term. The approach of [286] can be generalized to data potentials built from the combination of $L$ linear constancy constraints, that is,

$$\rho_{data}(x, I_1, I_2, w) = \sum_{\ell=1}^{L} \phi(A_\ell(x)^\top w(x) + B_\ell(x)). \tag{28}$$

For this kind of data potential, the regularisation tensor is defined by

$$\mathbf{R}_\rho(x) = \sum_{\ell=1}^{L} \mathbf{K}_\rho * A_\ell(x)A_\ell(x)^\top, \tag{29}$$

where $\mathbf{K}_\rho$ is a Gaussian kernel and $s_1, s_2$ are the eigenvectors of $\mathbf{R}_\rho(x)$. Taking $L = 1$, $A_1(x) = \nabla I(x)$ and $B_1(x) = I_t(x)$ yields the brightness constancy constraint, and the regularization tensor reduces to the structure tensor. For more elaborated data terms, as the combination of normalized brightness and gradient constancy used in [286], the resulting axes are more consistent with the data constraints.

### 5.1.2. Non-local regularization

The gradient of the flow can only provide a local constraint on the interaction between pixels. Assuming long range interactions can capture more precisely the form of the motion field. Such non-local regularization has been recently investigated in [79,154,231,269] by describing the structure of the flow in an extended neighborhood $\mathcal{N}(x)$ in a discrete setting as:

$$\rho_{reg}(x, w) = \sum_{y \in \mathcal{N}(x)} k(x, y, I_1)\, \phi\Big(\|w(x) - w(y)\|^2\Big). \tag{30}$$

The weights $k(x, y, I_1)$ indicate which pixel $y \in \mathcal{N}(x)$ should share a similar motion with pixel $x$. They are derived from the bilateral filter, favoring small distances in the spatial and color spaces [281]:

$$k(x, y, I_1) = \exp\left(-\frac{\|x - y\|^2}{\sigma_s^2} - \frac{\|I_1(x) - I_1(y)\|^2}{\sigma_c^2}\right), \tag{31}$$

where $\sigma_s$ and $\sigma_c$ control the influence of spatial distance and color similarity. This approach is image-driven in a similar way to local weighting (25), in the sense that the smoothness is weighted by the image edges. Nevertheless, the integration on a large neighborhood mitigates the influence of local gradients and more globally exhibits the structure of the objects. It is implemented as an alternate weighted median filtering in [231] and interpreted as a low-level soft segmentation in [272].

The high-order regularization causes severe optimization difficulties discussed in Section 5.2, in particular in terms of computational cost, increasing with the size of the neighborhood $\mathcal{N}(x)$.

### 5.1.3. Temporal coherence

A natural idea is to extend the spatial regularization described above to the temporal dimension, assuming that motion varies smoothly across consecutive frames. Similarly with the spatial case, smoothness on the time axis can be achieved either locally, based on the temporal gradient, or taking into account a longer interval by working on trajectories.

*Constraint on the temporal flow gradient.* The most straightforward way to model temporal smoothness is to penalize the temporal flow gradient, analogously with the spatial flow gradient in Section 5.1.1 [185]. This simple extension is however unrealistic since motion of objects necessarily implies a temporal change of flow fields. Thus, regularization on the spatio-temporal direction of the motion field is more appropriate, and is achieved in

[60,187,267,286] by extension of the spatial gradient penalties described in Section 5.1.1 to the spatio-temporal dimension. However, the performance of local temporal regularization is deceiving in most cases.

*Constraint on the trajectory*. Regularization based on temporal derivatives is unable to model large displacements. In this case, temporal coherence is more adequately modeled along the trajectories of objects. It was done in [29], who does not model explicitly the trajectories but estimates temporal changes on warped images to alleviate the problem of large displacements. However, the warping is done sequentially in the forward direction and is thus prone to propagate errors. In the same spirit, Volz et al. [253] considers the same coordinate system for groups of five frames, which implies an implicit natural registration. The estimation is done jointly in all frames of the sequence, which overcomes the lack of feedback with previous frames of [29]. The trajectory constraint in [93] is explicitly imposed by modeling the flow field as linear combination of long-term trajectory bases, obtained from few reliable tracks.

### 5.2. Optimization

As previously mentioned the optimization strategy employed to minimize (22) has a decisive influence on the final result. We give an overview of the main continuous and discrete optimization methods and point out their adaptability to specific energy terms.

#### 5.2.1. Continuous methods

*Resolution of the Euler–Lagrange equations*. The Euler–Lagrange equations give necessary conditions for minimizing energy of the form

$$E(w) = \int_\Omega F(x, w, \nabla w)\, dx, \tag{32}$$

which is the case of (22) with $F(x, w, \nabla w) = \rho_{data}(x, I_1, I_2, w) + \lambda \rho_{reg}(x, \nabla w)$ when the regularization is a function of the flow gradient (23). They provide the following system of partial differential equations:

$$\begin{cases} \dfrac{\partial \rho_{data}(x, I_1, I_2, w)}{\partial u} - \lambda \operatorname{div}\left(\dfrac{\partial \rho_{reg}(x, \nabla w)}{\partial \nabla u}\right) = 0 \\ \dfrac{\partial \rho_{data}(x, I_1, I_2, w)}{\partial v} - \lambda \operatorname{div}\left(\dfrac{\partial \rho_{reg}(x, \nabla w)}{\partial \nabla v}\right) = 0, \end{cases} \tag{33}$$

where $\frac{\partial}{\partial z}$ is the partial derivative operator with respect to $z$, $\operatorname{div}(\cdot)$ is the divergence operator. The system (33) can be rewritten by introducing the diffusion tensor $\mathbf{D}$ accounting for the relation with diffusion equations [266]:

$$\begin{cases} \dfrac{\partial \rho_{data}(x, I_1, I_2, w)}{\partial u} - \lambda \operatorname{div}\left(\mathbf{D}(x) \nabla u(x)\right) = 0 \\ \dfrac{\partial \rho_{data}(x, I_1, I_2, w)}{\partial v} - \lambda \operatorname{div}\left(\mathbf{D}(x) \nabla v(x)\right) = 0. \end{cases} \tag{34}$$

The analogy with diffusion equations makes explicit the direction and magnitude of the smoothing, which correspond respectively to the eigenvectors and eigenvalues of $\mathbf{D}(x)$.

The discretization of the gradient and divergence operators yields a large system of equations to be solved. If the system is linear, its sparsity makes it well suited for iterative solvers like Gauss–Seidel or successive over-relaxation (SOR) [45]. Nevertheless, the linear case is in practice only encountered in the model of [120] using quadratic penalties and a linearized data constraint. To cope with non-linearity, the typical approach [45,265] is to resort to time-lagged schemes [62] by handling each source of non-linearity with fixed point iterations, turning the problem into a succession of linear systems, and updating iteratively the non-linear parts. Convergence of the scheme is ensured if the linear systems are solved exactly, but approximations and frequent updates often yield in practice good results and much faster convergence.

Fast computational schemes have been employed for achieving near real-time performance. A multigrid framework has been proposed in [50]. The scheme is very efficient, but it is problem-specific and requires a substantial implementation effort. Another approach is to consider the solution of the Euler–Lagrange equation as the steady state of the corresponding diffusion process (34), and use the Fast Explicit Diffusion (FED) principle [105] to accelerate convergence by adapting the time steps. The implementation of [107] exploits the natural parallelization of explicit schemes to achieve a quasi real-time version of the variational method of [286] on GPU, based on FED.

The approach of [45] has become standard because of its simplicity, the wide range of models it can handle, and its good experimental performances even for non-convex energies [45,164,253].

*Early discretization*. The computation of the Euler–Lagrange equations can be complex or impossible for some forms of the energy (22). Moreover, one can argue that the discretization of the Euler–Lagrange equations can generate numerical errors with respect to the original energy [204]. A way to alleviate these shortcomings is to avoid computing the Euler–Lagrange equations and instead to directly discretize the energy (22). The equation to solve is then simply the cancellation of the differentiated discretized energy:

$$\frac{\partial E(w)}{\partial w} = 0. \tag{35}$$

Solving (35) amounts to inverting a large linear system, similar to the one obtained by Euler–Lagrange equation discretization for energies of the form (32). Employing a fixed-point iterations scheme to cope with non-linearity amounts to an IRLS approach, which is shown by [165] to be equivalent to the above described resolution of the Euler–Lagrange equations with fixed-point.

Contrary to the Euler–Lagrange scheme, the regularization is not imposed to be a derivative of $w$, and non-local regularization terms (30) can be handled. However, such an approach yields a dense linear system, not solvable with standard iterative methods. In [154], a linear-time method is proposed to compute matrix product by successive Gaussian filtering, allowing to efficiently inverse the dense linear system with a conjugate gradient solver. The method of [232] also exploited the ability of handling more general energies to optimize learned data and regularization terms. The general experimental study of [231] and the complex method of [233] also follows this approach, with Graduated Non Convexity (GNC) to cope with multimodality of the energy [27,33].

Instead of solving (35), a gradient descent method can be applied to minimize the discretized energy. Due to the large scale of the problem, Newton methods requiring the inversion of the Hessian of the energy are not applicable, and only quasi-Newton methods are computationally tractable. A few works have explored this direction, with Truncated Newton [134,135] or L-BFGS [75].

*Half-quadratic minimization*. Instead of solving directly the energy (22), a number of optimization methods proceed to the addition of an auxiliary variable splitting the original problem into easier sub-problems. The half-quadratic minimization falls in this class. It can be shown that under non-restrictive conditions [58,96], a function $\phi$ can be rewritten as the following function $\Phi$ introducing the additional variable $\gamma \in ]0, 1]$:

$$\phi(z) = \inf_\gamma \Phi(z, \gamma) = \inf_\gamma \left(\gamma z^2 + \psi(\gamma)\right), \tag{36}$$

where the function $\psi$ can be explicitly derived from $\phi$. The robust non-convex penalty $\phi(z)$ in data and regularization potentials can be replaced by $\Phi(z, \gamma)$ in the data and regularization terms, so that

the optimization in $z$ becomes easy since it involves only quadratic terms. Moreover, the minimization w.r.t. $\gamma$ has a closed-form solution:

$$\widehat{\gamma} = \arg\min_{\gamma} \Phi(z,\gamma) = \frac{\phi'(z)}{2z}. \tag{37}$$

The original minimization problem in $z$ is thus turned into alternate simple optimizations on $z$ and $\gamma$. This approach also leads to the IRLS algorithm described in Section 4.2.

The introduction of this approach for optical flow estimation coincided with the use of robust penalties for discontinuity preserving regularization [7,32,77,174] and was more recently exploited in [65,112].

*Proximal splitting*. Another successful optimization method based on alternate minimization of simple sub-problems has been proposed by [8] and used for optical flow in [285]. The data and regularization terms are splitted and associated to separate variables, which are quadratically coupled by a third term:

$$\begin{aligned} E_{\text{split}}(w,v) = & \underbrace{\int_\Omega \rho_{data}(x,I_1,I_2,w)dx}_{E_{data}} + \frac{1}{2\varepsilon}\int_\Omega \|w(x)-v(x)\|^2 dx \\ & + \lambda \underbrace{\int_\Omega \rho_{reg}(x,v)dx}_{E_{reg}}, \end{aligned} \tag{38}$$

where $v$ is an auxiliary variable. The parameter $\varepsilon$ sets the intensity of the coupling. If $\varepsilon$ is small, (38) tends to the original univariate problem (22). The minimization w.r.t. each variable is the computation of the proximal operator of $E_{data}$ and $E_{reg}$:

$$\begin{cases} \arg\min_w E_{\text{split}}(w,\widehat{v}) = \text{prox}_{E_{data}}(\widehat{v}) \\ \arg\min_v E_{\text{split}}(\widehat{w},v) = \text{prox}_{E_{reg}}(\widehat{w}), \end{cases} \tag{39}$$

where the proximity operator of a function $f$ is defined by

$$\text{prox}_f(z) = \arg\min_u \left( f(u) + \frac{1}{2}\|u-z\|^2 \right). \tag{40}$$

The minimization problems (39) can be viewed as alternating coarse pixel-wise data matching and denoising of the flow field. A large number of variants and generalizations of this proximal splitting idea exist, among which the alternating direction method of multipliers (ADMM) [38] or the formalization in a primal–dual framework described in [55].

This approach has initially been designed for $L_1$ penalty on the data and regularization terms, a.k.a. *TV-*$L_1$ model [8,285], for the simplicity of the resulting optimization sub-problems. The optimization of the data term with $v$ fixed is efficiently performed by a thresholding scheme [280], and fixing $w$ yields the Rudin–Osher–Fatemi model [210], optimized with the duality based algorithm of [54]. An advantage is that a differentiable approximation of the $L_1$ norm is not required, as in [45], and one can solve for the exact *TV*–$L_1$ model.

In a general point of view, the independence of the optimization of data and regularization parts allows one to design dedicated minimization schemes in a variety of cases. The restriction is to be able to compute the proximal operators, and the convergence is ensured only for convex energies. For the data part, the thresholding scheme of [285] is applicable in some cases for the $L_1$ norm, and efficient solutions have been found to handle an additional fundamental matrix constraint [262], a truncated $L_1$ norm of normalized cross correlation [269], or mutual information [196]. Another advantage of the pixel-wise nature of the data part minimization is that it enables parallel implementation strategies which can dramatically speed up the algorithm and reach real-time [285]. Based on the pixel-wise property, [227] proposes a discrete exhaustive matching for optimization in $w$, which opens the usage of arbitrary data terms, only limited by the computational cost of the matching. Patch-based similarity measures have also recently been implemented with the fast PatchMatch algorithm [15] by [114].

Concerning the regularization part, it is possible to minimize non-local regularization terms [79,272]. From an evaluation viewpoint, the decoupling in the minimization process also allows for a fair comparison of the effects of different data and regularization terms [252].

#### 5.2.2. Combinatorial methods

In a discrete setting, the solution of the minimization of (22) is searched in a set of discrete labels $\mathcal{L}$ corresponding to a quantization of the continuous motion vector space or the selection of a finite subset of motion vectors. We give a fast overview of basic principles of optimization methods for this combinatorial problem. For a more complete analysis, see the recent review of [255]. The spatial discretization of (22) yields an energy in the Markov Random Field (MRF) framework:

$$E_D(w) = \sum_{x\in\Omega_D} \left[ \rho_{data}(x,I_1,I_2,w) + \sum_{y\in\mathcal{N}(x)} \rho_{MRF}(x,y,w) \right], \tag{41}$$

where $\Omega_D$ is the discrete image domain, $\mathcal{N}(x)$ denotes the pixels interacting with $x$ in the model and $\rho_{MRF}$ is the discrete version of $\rho_{reg}(x,y,I_1,I_2,w)$ which explicitly takes into account the interaction between two neighboring pixels $x$ and $y$.

An important advantage of discrete optimization over the continuous approach is that it does not require differentiation of the energy and can thus handle a wider variety of data and regularization terms. On the counterpart, a trade-off has to be found between the accuracy of the motion labeling and the size of the search space $\mathcal{L}$. Indeed, discrete optimization methods are usually severely limited in terms of accuracy and speed by the number of labels, particularly for optical flow where subpixel accuracy is necessary, and where the two-dimensional motion space becomes more rapidly intractable than the one-dimensional stereo case for instance. Therefore, the design of the label space $\mathcal{L}$ is a crucial component of discrete optimization methods for optical flow.

Among early methods for minimizing (41), simulated annealing [97] offers proved convergence towards the global minimum based on stochastic relaxation. However, the optimal convergence is guaranteed only under prohibitive computational cost. An approximate solution can be obtained much faster with the Iterated Conditional Modes [25] or its High Confidence First variant [61], operating by iterative local minimizations of the energy, but this local optimum often yields poor results compared to modern methods [237], especially for non-convex functions. A multi-scale MRF approach was designed in [117] where the energy function is minimized over nested subspaces of the original discrete label space. This constrained optimization scheme offered simple implementation and fast convergence towards high quality estimates for non-linear problems such as motion computation.

*Graph cut*. The work of [39] gave rise to rapidly growing research interest and advances on graph cut approaches for MRF-based energy minimization. The basis of graph cuts is the max-flow/min-cut problem consisting in finding the optimal path between two nodes in a directed graph, solvable by many algorithm in polynomial time [91,102]. It is possible to model an undirected MRF structure as a directed graph by introducing two additional *source* and *sink* nodes, and then interpret the min-cut partition as a binary label segmentation of the MRF energy. The global minimum of the binary optimization can be guaranteed for pairwise interactions and submodular functions. In summary,

it is possible to find the global optimum for the energy (41) under the following conditions:

- submodularity of $\rho_{MRF}(x, y, w)$,
- pairwise interactions,
- binary labels.

Research on graph cuts has attempted to overcome these three constraints, based on the original max-flow/min-cut algorithm.

Submodularity of the pairwise term is a required property for convergence of the algorithm. Finding good approximate solution can be achieved with the Quadratic Pseudo Boolean Optimization (QPBO) [35,148,209], leading to an optimal but partial labeling. Dealing with higher-order local interactions between pixels has also been addressed in several recent works [86,128,144,150].

Submodularity and pairwise interactions are not too restrictive constraints for a large range of computer vision problems and let room for a number of applications. On the contrary, the cardinality of the label space satisfies $|\mathcal{L}| > 2$ in most cases, so that the binary label requirement is a much harder limitation. The extension of binary graph cuts to multi-label have mostly been realized through iterative *move-making* techniques. The idea is to create at each iteration a binary-labeled space composed of the current solution and a new proposal label. The label set $\mathcal{L}$ is thus explored progressively by each new proposal. If each binary minimization performed with the max-flow/min-cut method is ensured to be optimal, a decreasing of the energy is guaranteed at each iteration, and the method converges to a local minimum. It can be noticed that the class of *move-making* methods is not restricted to graph cuts and the *moves* can be achieved by other techniques such as the variational approach of [242], optimizing in the continuous space by relaxing the binary variable.

The elements of a *move-making* method are the *move-space*, specifying the set of new labeling proposals at each iteration, and the way the *moves* themselves are performed. The most popular *move-space* is the *expansion-move*, where the proposal labeling is defined as a constant label map. Another common alternative is the $\alpha\beta$-swap move based on a label exchange at pixels having labels $\alpha$ or $\beta$. The *range-move* [250] extends binary proposal choice to a larger range of labels, in the case of ordered label spaces. Precomputed labelings computed with independent and possibly continuous methods, can also serve as proposals [161].

Additionally, computational efficiency of graph-cut approaches have been addressed by several works. The most representative ones are [152], operating in a Primal–Dual framework and speeding up convergence by minimizing the Primal–Dual gap, and on the other hand, [145] working on dynamic MRF and exploiting previous iterations as initializations. For more details about existing *move-spaces*, see [249,255]. Applications for optical flow have increased rapidly in recent years [59,64,88,101,99,161,164,234,279].

*Message passing*. Belief propagation is based on the *max-product* algorithm, which is able to find the MAP of a probability distribution expressed as a product of factors, and thus representable in a factor graph (see [157] for a detailed introduction). Taking the negative logarithmic version of such distribution amounts to considering MRFs of the form (41), which can motivate to rename the algorithm *min-sum* in this case. In a nutshell, the minimization is done by iteratively updating local *messages* reflecting influence of local label configurations on the energy. After convergence of the *messages*, they can be used to define the probability of assigning a given label to a node, and the label with the maximum probability is chosen. The *max-product* has originally been designed for tree structures and is guaranteed to find the global optimum in this case. Nevertheless, it can also be used for MRFs exhibiting cycles (it is referred to as loopy belief propagation in this case [200]), without convergence guarantees but showing good experimental results in a large number of computer vision problems [136,236]. The Tree Reweighted message passing approach [254] deals with similar concepts, but with a particular message passing strategy based on tree representations. The sequential approach of [147] has proven to yield good results and computational performance compared to other discrete methods in [237]. It has been applied for optical flow estimation in [104,121,160,181]. It is also exploited for image matching involving large deformation models to cope with high computational complexity, as described in [216] or in the SIFT-Flow algorithm [166].

## 6. Occlusion handling

The data costs described in Section 3 have been defined as measures of some data constancy between two corresponding points. However, they do not take into account the case of pixels without correspondence, namely occlusions. The occlusion issue in motion estimation, stereovision or image correspondence has motivated different types of investigation [10,116,34,129,149,162,225,273,276]. Occluded regions in the current image of the video sequence are defined as the set of pixels which become hidden by occluding (moving) objects in the next image, as illustrated in Fig. 3. These pixels have no corresponding points in the next image and motion is *not observable* at their locations. Points on (static or moving) objects which leave the camera field of view to the next image are part of a similar situation. Hence, the occlusion issue must be carefully addressed to ensure a reliable and accurate computed optical flow. This is particularly the case for large displacements [229], as in the MPI Sintel dataset [52]. Most methods implicitly deal with occlusion, together with other sources of violation of the data constancy assumption, by using robust norms in the data term, as emphasized in Section 3. To describe explicit occlusion handling approaches, we must further distinguish between *occlusion detection*, i.e., segmenting the image into occluded and non-occluded regions, and *occlusion filling*, i.e., recovering the missing flow subfields in occluded regions.

### 6.1. Occlusion detection

Occlusion detection has been mostly investigated as a subsequent operation to motion computation, by thresholding a consistency measure issued from the estimated motion field. Several criteria have been proposed as geometric forward–backward motion mismatch [127,181], mapping unicity [279], divergence of the flow field [14] or data constancy violation [276]. The main limitation of the sequential approach is that accuracy of occlusion detection is highly dependent on the quality of the motion estimation. Several flow and image criteria have also been combined in a learning framework [126].

Other approaches estimate the occlusion map jointly with the motion field in an alternate optimization scheme [10,88,127,197,244]. In the continuous setting of [10] involving a sparsity constraint for occlusion detection, the data constancy deviation is balanced by an estimated continuous residual intensity field, from which occluded points are retrieved by thresholding. The usual drawback of alternate optimization schemes is the convergence towards local minima due to the strong coupling between motion and occlusions. This problem is addressed in [88] by introducing an occlusion confidence map in the model, guiding occlusion estimation at each iteration independently from the current motion estimate.

Finally, occlusion map can also be retrieved from the decomposition of the motion field in layers [223,233,234]. Indeed, occluded

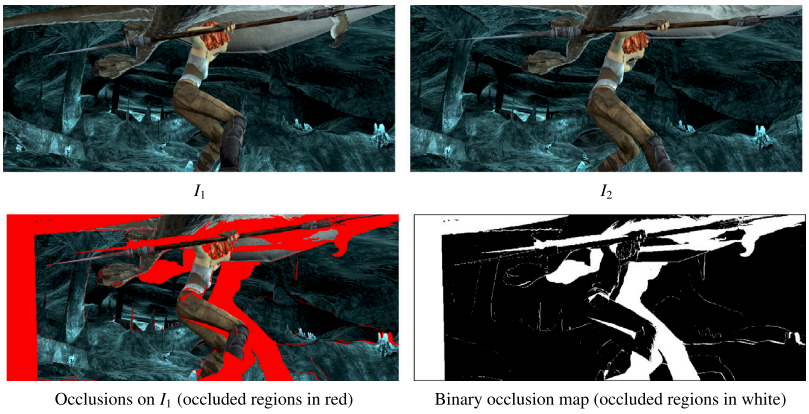


**Fig. 3.** Example of occluded regions.

parts of each layers can be naturally derived from their depth ordering and the way they overlap with each other.

### 6.2. Occlusion filling

When occluded regions are known, occlusion filling with motion vectors is conceptually closely related to image inpainting, since it recovers motion in regions where it is by definition not observable. Inpainting methods can be coarsely divided into two classes, diffusion-based methods [24,56] and exemplar-based methods [6,68,151]. A synthesis of these two approaches has been investigated in [51] in a variational framework.

Let us denote $\mathcal{O} \subset \Omega$ the set of occluded pixels of $I_1$. Classical methods for occlusion filling with motion vectors operate in a global regularization framework by cancelling the data term at occluded points, that is $\rho_{data}(x_o, w, I_1, I_2) = 0, \forall x_o \in \mathcal{O}$. Thus, this class of diffusion-based (or geometry-oriented) occlusion filling relies on the diffusion process of the regularization to propagate motion from non-occluded regions to occluded regions via partial derivative equation (PDE) resolution [10,197,279].

It is known that diffusion-based inpainting methods perform well in case of thin missing areas or cartoon-like images, but they are usually unable to handle for large missing regions. Rather than simply cancelling the data term, it is then necessary to design valid data costs for occluded pixels when dealing with large occlusion regions. This is achieved in [88], following the inpainting analogy and inspired by the superior performances of exemplar-based approach for large missing regions. At occluded pixels $x \in \mathcal{O}$, the data cost imposes proximity with the motion vector of a matched non-occluded pixel $m(x_o) \in \Omega \setminus \mathcal{O}$, assumed to belong to the same object as $x_o$ and thus to have similar motion:

$$\rho_{data}(x_o, w, I_1, I_2) = \|w(x_o) - w(m(x_o))\|^2. \tag{42}$$

Another possible definition of data cost at occluded pixels can be found in [14], under the assumption of temporal constancy of motion. Assuming that a pixel $x_o \in \mathcal{O}$ of $I_1$, occluded in the next frame $I_2$, is non-occluded in the previous frame $I_0$, backward data constancy is imposed at occluded pixels:

$$\rho_{data}(x_o, w, I_1, I_2) = \phi(I_0 - I_1(x_o - w(x_o))). \tag{43}$$

## 7. Combining feature matching and optical flow

An increasing challenge for optical flow estimation is to handle very large displacements and deformations, as it is reflected in the recent MPI Sintel benchmarks [52], where is it not rare to find displacements of more than 100 pixels. As a consequence, the gap between feature matching and optical flow tends to vanish, and several methods have tried to combine density and accuracy of optical flow with the ability to capture large displacements of feature matching.

Finding correspondences by matching image features can be considered as a *local* parametric approach since a given neighborhood of the pixel to match is assumed to translate or undergo another parametric transformation towards its correspondence location. The similarity measures can be patch-based distances (see Section 3.1.2), or more complex and sparse feature descriptions often based on histogram of oriented gradients [20,71,169,282], or segment matching [258]. The fundamental difference lies in the optimization process, since the parametric formulation has a differential optimization process imposing linearization, whereas feature matching explores a discrete space of admissible correspondences. Although being integer displacements and prone to errors, feature matching can thus handle large displacements without coarse-to-fine-schemes, with arbitrarily complex similarity measures. Regarding their complementarity of advantages, the combination of feature matching with regularization approaches is therefore of upmost interest.

We consider three approaches to obtain dense and accurate motion fields with feature matching: local filtering of correspondences, integration in a variational framework, and generation of coarse initialization for variational refinement.

### 7.1. Correspondence filtering

Pure feature matching has long been considered unable to produce dense flow fields with competitive accuracy with the previously described *local* and *global* approaches. There are three reasons for this:

1. The optimization of the similarity measure is often performed with exhaustive search, which induces prohibitive computational cost.
2. Repetitive textures or uninformative regions are sources of ambiguities for the matching process and generate large errors.
3. The correspondence process usually limits the accuracy to integer displacements, contrary to *local* and *global* approaches working in the continuous $\mathbb{R}^2$ space.

Several recent advances attempted to overcome these three seemingly inherent limitations.

Research on speeding up block matching include multi-scale search strategies [243], integral images [83] or search in trees

[158], but the recent most spectacular contribution was achieved in [15,16] with the PatchMatch algorithm. The method scans the image in the lexicographic and inverse order, and alternates two simple steps at each pixel: the *propagation* step minimizes locally the data cost in the space composed by the current pixel and its two predecessors in the scanning order, the second step proposes a small set of new candidates randomly chosen in the neighborhood of the current correspondence. It is easy to extend the matching to more complex transformations than translation, like rotation or scale factor [16] by increasing the degree of the search space. The method was originally designed for image editing and was applied to several other applications [16], with impressive results regarding the low computation time. For motion estimation, the interesting property is that the computational cost is not affected by the spatial extension of the search space, so that no trade-off has to be found between speed and displacement range.

The problem of matching ambiguities comes from the lack of discriminative power of the data cost. Without resorting to explicit regularization (Section 5), the coherency induced by simple local filtering of patch correspondences [122] has been shown to be sufficient in practice to reduce most ambiguities and provide satisfactory dense results. The filtering is achieved in [122] by guided filtering [111] and in [172] by weighted median.

Subpixel accuracy is usually reached by upscaling image resolution or by locally approximating the matching surface by a polynomial function. The induced computational cost is reduced in [122,227] by GPU implementation, and can also be handled by iterative refinement [160].

The combination of the three ingredients has led to the development of competitive optical flow estimation methods based on pure feature matching locally filtered [122,172,238].

### 7.2. Feature matching in global regularized model

As discussed in Section 2.4, a major limitation of the global variational framework is the loss of small and fast objects due to the use of a coarse-to-fine estimation scheme. One recently investigated approach to overcome this problem is to integrate an information from feature matching into the variational framework, thus combining advantages of both methods.

The approach of [47], inspired by [113], made the first step in this way by adding to the classical data potential a new constraint taking into account an off-line computation of sparse feature correspondences. Let us denote $w_c$ the displacement field obtained with a possibly sparse feature matching process. The new combined data potential is then

$$\rho^{+}_{data}(x, I_1, I_2, w, w_c) = \rho_{data}(x, I_1, I_2, w) + \beta \rho_{corresp}(x, w, w_c), \tag{44}$$

where $\beta$ is a trade-off parameter and the matching potential is defined as:

$$\rho_{corresp}(x, w, w_c) = \delta(x, w_c)\, c(x, w_c)\, \phi(\|w(x) - w_c(x)\|^2). \tag{45}$$

The binary function $\delta(x, w_c)$ returns 1 if $w_c(x)$ is defined at $x$ and 0 otherwise, and the weights $c(x, w_c)$ correspond to the matching cost. The third term of (45) imposes the motion field to be close to the motion vectors obtained by feature matching.

The advantage of this approach is that the term $\rho_{corresp}(x, w, w_c)$ is both differentiable and valid for large displacements. Problems related to the use of coarse-to-fine schemes are thus avoided. The main drawback is that the importance of the matching term $\rho_{corresp}(x, w_c, w)$ relatively to the data term $\rho_{data}(x, I_1, I_2, w)$ is mostly determined by the value of the matching cost $c(x, w_c)$. Consequently the final variational estimation is extremely dependent on the reliability of the confidence measure, which is very difficult to guarantee as emphasized in [42]. Matching errors are thus easily driven by $c(x, w_c)$ and are likely to have a high impact on the final result. Reducing the impact of local feature matching errors by the regularization and the robust penalization $\phi(\cdot)$ is insufficient in practice in a lot of cases.

Recently, the authors of [268] based their method on [47] by taking the model (45) and improving the feature matching stage. They demonstrated a significant performance improvement due to the increased reliability of the matching. The method of [42] is also built upon [47] with a modified matching component by using segment matching [258]. The matching term is generalized to handle weakly localized line features. It is done through a point-to-line distance in addition to the point-to-point distance of (45), combined with a confidence measure for segment matching based on the assumption of a linear mapping between segments. Moreover, considering the unreliability of confidence measure, [42] gets rid of the matching cost $c(x, w_c)$ and relies only on the regularization for robustness to matching errors. To sum up, the tendency is to take into account the great dependency of the estimation on the quality of the feature matching, and thus to concentrate most efforts on the design of robust matching algorithms.

### 7.3. Feature matching for coarse initialization

Another recently investigated approach taking advantage of feature matching consists in exploiting the (possibly sparse) field of integer displacements to provide a coarse but relevant initialization for a variational refinement [59,181,279]. These methods are composed of three steps:

1. A feature matching step provides a limited number of candidates at each pixel.
2. These candidates serve as labels for the discrete optimization (see Section 5.2) of a global regularized model.
3. The resulting coarse optical flow field is refined with one of the variational optimization techniques described in Section 5.2.

The idea is that steps 1 and 2 handle indifferently small and large displacements of objects at any scale, and the initialization is assumed to be good enough to avoid the coarse-to-fine scheme in step 3. The specificity of [279] is that the three steps are repeated at each level of a multiresolution pyramid. Since steps 2 and 3 have already been discussed in previous subsections, we focus on the specificity of step 1, that is, the production of candidates from feature matching.

In [279], the feature matching is only performed at a restricted number of keypoints. After pruning of similar vectors, only a few displacement vectors are retained. Each of these vectors is expanded to produce a global constant flow field, used as candidates for step 2.

The approach of [181] is to consider an integer discretization of the two-dimensional motion field. Based on the observation that correlation-based patch matching is able to reproduce coarsely the motion distribution pattern of ground truth motion fields, the discretized motion space is delineated by the matching vectors.

The approach of [59] is close to [181] since the idea is to rely on the dominant motion patterns of dense patch matching. In [59], the candidates to feed discrete optimization are obtained by explicitly clustering PatchMatch motion fields [16] to keep only dominant patterns.

Another strategy has recently been investigated in [88], exploiting the aggregation paradigm. Feature matching is used as a part of a generation procedure of continuous motion candidates, and the best candidates are then selected in the aggregation step with a global model, without need of subsequent variational refinement.

## 8. Conclusion and perspectives

Through the analysis of optical flow literature, we have introduced principles of optical flow modeling and computation, and classified the main methodological aspects of existing methods. From this taxonomy, interests and limitations of each class have been put forward.

This study also exhibits the main research directions for improving current state-of-the-art. In terms of performance, it can be considered from results of the Middlebury benchmark [13] that for a range of sequences with small displacements, moderate intensity changes and piecewise smooth displacements of large structures, globally regularized models and joint estimation and segmentation methods are able to obtain satisfying results. However, moving away from these dataset characteristics, as it is the case in recent benchmarks [52,94] or applicative domains [89], unveils the shortcomings of current methods in challenging situations, and it opens the way to potentially large improvements.

In particular, some of the main difficulties are caused by large displacements. To avoid the drawbacks of coarse-to-fine schemes of variational methods, an active research direction focuses on the design of computationally tractable discrete optimization methods for the large scale problem of optical flow. We have noticed that dealing with larger displacements brings optical flow closer to feature matching. However, attempts to combine these two worlds still consider them separately and try to fuse their independent results, often leading to high sensitivity to matching errors. Another essential problem lies in large occlusions that are inevitably associated to large motion. Large intensity changes appearing in real environments or in specific applications, or caused by large deformations, is another main problem of current methods. The most promising direction seems to be the spatial adaptation of the data term.

Finally, another remaining challenge is to reduce the computational cost of the best performing methods.

We observe an evolution of the scope of optical flow, which was originally devoted to image sequences with high frame-rate, for which deviations from small displacements and brightness conservation stayed reasonably low. As this well identified problem becomes better handled, optical flow estimation tends to apply on larger class of sequences. Thus, the gap between optical flow and other correspondence problems like registration or feature matching progressively reduces and opens the way to unified correspondence methods.

## Acknowledgments

This work was realized as part of the Quaero program, funded by OSEO, French State agency for innovation. It was also partly supported by the France-BioImaging project granted by the Investissement d'Avenir program, and by UMR 144 CNRS of Institut Curie.

## References

[1] S. Ali, M. Shah, Human action recognition in videos using kinematic features and multiple instance learning, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2) (2010) 288–303.
[2] L. Alvarez, J. Esclarín, M. Lefebure, S. Javier. A PDE model for computing the optical flow, in: Proc. XVI Congreso de ecuaciones diferenciales y aplicaciones, Las Palmas de Gran Canaria, Spain, September 1999, pp. 1349–1356.
[3] F. Amat, E.W. Myers, P.J. Keller, Fast and robust optical flow for time-lapse microscopy using super-voxels, Bioinformatics 29 (3) (2013) 373–380.
[4] T. Amiaz, N. Kiryati, Piecewise-smooth dense optical flow via level sets, Int. J. Comput. Vision 68 (2) (2006) 111–124.
[5] T. Amiaz, E. Lubetzky, N. Kiryati, Coarse to over-fine optical flow estimation, Pattern Recogn. 40 (9) (2007) 2496–2503.
[6] P. Arias, G. Facciolo, V. Caselles, G. Sapiro, A variational framework for exemplar-based image inpainting, Int. J. Comput. Vision 93 (3) (2011) 319–347.
[7] G. Aubert, R. Deriche, P. Kornprobst, Computing optical flow via variational techniques, SIAM J. Appl. Math. 60 (1) (1999) 156–182.
[8] J.-F. Aujol, G. Gilboa, T. Chan, S. Osher, Structure-texture image decomposition – modeling, algorithms, and parameter selection, Int. J. Comput. Vision 67 (1) (2006) 111–136.
[9] S. Ayer, H. Sawhney. Layered representation of motion video using robust maximum-likelihood estimation of mixture models and MDL encoding, in: IEEE Int. Conference on Computer Vision (ICCV), Boston, June 1995, pp. 777–784.
[10] A. Ayvaci, M. Raptis, S. Soatto, Sparse occlusion detection with optical flow, Int. J. Comput. Vision 97 (3) (2012) 322–338.
[11] S. Baker, M.J. Black, J. Lewis, S. Roth, D. Scharstein, and R. Szeliski. A database and evaluation methodology for optical flow, in: International Conference on Computer Vision (ICCV), Rio de Janeiro, Brazil, October 2007, pp. 1–8.
[12] S. Baker, I. Matthews, Lucas–Kanade 20 years on: a unifying framework, Int. J. Comput. Vision 56 (3) (2004) 221–255.
[13] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, R. Szeliski, A database and evaluation methodology for optical flow, Int. J. Comput. Vision 92 (1) (2011) 1–31.
[14] C. Ballester, L. Garrido, V. Lazcano, V. Caselles, A TV-L1 optical flow method with occlusion detection, Pattern Recogn. (2012) 31–40.
[15] C. Barnes, E. Shechtman, A. Finkelstein, D.B. Goldman, Patchmatch: a randomized correspondence algorithm for structural image editing, ACM Trans. on Graphics, vol. 28, ACM, 2009, p. 24.
[16] C. Barnes, E. Shechtman, D.B. Goldman, A. Finkelstein, The generalized patchmatch correspondence algorithm, in: European Conference on Computer Vision (ECCV), 2010, pp. 29–43.
[17] J. Barron, D. Fleet, S. Beauchemin, Evaluation of optical flow, Int. J. Comput. Vision 12 (1) (1994) 43–77.
[18] T. Basha, Y. Moses, N. Kiryati, Multi-view scene flow estimation: a view centered variational approach, Int. J. Comput. Vision 101 (1) (2013) 6–21.
[19] A. Basset, P. Bouthemy, C. Kervrann, Recovery of motion patterns and dominant paths in videos of crowded scenes, in: International Conference on Image Processing (ICIP), Paris, October 2014.
[20] H. Bay, T. Tuytelaars, L. Van Gool, Surf: speeded up robust features, in: European Conference Computer Vision (ECCV), 2006, pp. 404–417.
[21] S.S. Beauchemin, J.L. Barron, The computation of optical flow, ACM Comput. Surv. (CSUR) 27 (3) (1995) 433–466.
[22] F. Becker, B. Wieneke, S. Petra, A. Schroder, C. Schnörr, Variational adaptive correlation method for flow estimation, IEEE Trans. Image Process. 21 (6) (2012) 3053–3065.
[23] J.R. Bergen, P. Anandan, K.J. Hanna, R. Hingorani, Hierarchical model-based motion estimation, in: European Conference on Computer Vision (ECCV), Springer, 1992, pp. 237–252.
[24] M. Bertalmio, G. Sapiro, V. Caselles, C. Ballester, Image inpainting, in: Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques, 2000, pp. 417–424.
[25] J. Besag, On the statistical analysis of dirty pictures, J. Roy. Stat. Soc. Ser. B (Methodol.) (1986) 59–302.
[26] J. Bigun, G.H. Granlund, J. Wiklund, Multidimensional orientation estimation with applications to texture analysis and optical flow, IEEE Trans. Pattern Anal. Mach. Intell. 13 (8) (1991) 775–790.
[27] M. Black, P. Anandan, The robust estimation of multiple motions: parametric and piecewise-smooth flow fields, Comput. Vis. Image Underst. 63 (1) (1996) 75–104.
[28] M. Black, A. Jepson, Estimating optical flow in segmented images using variable-order parametric models with local deformations, IEEE Trans. Pattern Anal. Mach. Intell. 18 (10) (1996) 972–986.
[29] M.J. Black, Recursive non-linear estimation of discontinuous flow fields, in: European Conference on Computer Vision (ECCV), 1994, pp. 138–145.
[30] M.J. Black, P. Anandan, A framework for the robust estimation of optical flow, in: International Conference on Computer Vision (ICCV), 1993, pp. 231–236.
[31] M.J. Black, Y. Yacoob, Recognizing facial expressions in image sequences using local parameterized models of image motion, Int. J. Comput. Vision 25 (1) (1997) 23–48.
[32] M.J. Black, Y. Yacoob, A.D. Jepson, D.J. Fleet, Learning parameterized models of image motion, in: Computer Vision and Pattern Recognition (CVPR), 1997, pp. 561–567.
[33] A. Blake, A. Zisserman, Visual Reconstruction, MIT press, Cambridge, 1987.
[34] M. Bleyer, C. Rhemann, M. Gelautz, Segmentation-based motion with occlusions using graph-cut optimization, in: DAGM Symposium on Pattern Recognition, Berlin, Germany, September 2006, pp. 465–474.
[35] E. Boros, P. Hammer, X. Sun, Network Flows and Minimization of Quadratic Pseudo-boolean Functions, Technical Report, Technical Report RRR 17-1991, RUTCOR, 1991.
[36] J.-Y. Bouguet, Pyramidal Implementation of the Lucas-Kanade Feature Tracker, Description of the Algorithm, Technical Report, Intel Microprocessor Research Labs, 1999.
[37] P. Bouthemy, E. François, Motion segmentation and qualitative dynamic scene analysis from an image sequence, Int. J. Comput. Vision 10 (2) (1993) 157–182.
[38] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends® Mach. Learn. 3 (1) (2011) 1–122.
[39] Y. Boykov, O. Veksler, R. Zabih, Markov random fields with efficient approximations, in: Computer Vision and Pattern Recognition (CVPR), 1998, pp. 648–655.

[40] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Trans. Pattern Anal. Mach. Intell. 23 (11) (2001) 1222–1239.
[41] G. Bradski, Opencv, Dr. Dobb's J. Softw. Tools (2000).
[42] J. Braux-Zin, R. Dupont, A. Bartoli, A general dense image matching framework combining direct and feature-based costs, in: International Conference on Computer Vision (ICCV), 2013.
[43] K. Bredies, K. Kunisch, T. Pock, Total generalized variation, SIAM J. Imag. Sci. 3 (3) (2010) 492–526.
[44] L.G. Brown, A survey of image registration techniques, ACM Comput. Surv. (CSUR) 24 (4) (1992) 325–376.
[45] T. Brox, A. Bruhn, N. Papenberg, J. Weickert, High accuracy optical flow estimation based on a theory for warping, in: European Conference on Computer Vision (ECCV), Prague, Czech Republic, May 2004, pp. 25–36.
[46] T. Brox, A. Bruhn, J. Weickert, Variational motion segmentation with level sets, in: European Conference on Computer Vision (ECCV), 2006, pp. 471–483.
[47] T. Brox, J. Malik, Large displacement optical flow: descriptor matching in variational motion estimation, IEEE Trans. Pattern Anal. Mach. Intell. 33 (3) (2011) 500–513.
[48] T. Brox, J. Weickert, B. Burgeth, P. Mrázek, Nonlinear structure tensors, Image Vis. Comput. 24 (1) (2006) 41–55.
[49] A. Bruhn, J. Weickert, C. Schnörr, Lucas/Kanade meets horn/schunck: combining local and global optic flow methods, Int. J. Comput. Vision 61 (3) (2005) 211–231.
[50] A. Bruhn, W. Weickert, A confidence measure for variational optic flow methods, Geom. Properties Incomplete Data (2006) 283–298.
[51] A. Bugeau, M. Bertalmío, V. Caselles, G. Sapiro, A comprehensive framework for image inpainting, IEEE Trans. Image Process. 19 (10) (2010) 2634–2645.
[52] D.J. Butler, J. Wulff, G.B. Stanley, M.J. Black, A naturalistic open source movie for optical flow evaluation, in: European Conference on Computer Vision (ECCV), Springer-Verlag, 2012, pp. 611–625.
[53] X. Cai, J.H. Fitschen, M. Nikolova, G. Steidl, M. Storath, Disparity and optical flow partitioning using extended Potts priors, Inform. Inference: J. IMA (2014) (submitted for publication) arXiv:1405.1594.
[54] A. Chambolle, An algorithm for total variation minimization and applications, J. Math. Imag. Vision 20 (1–2) (2004) 89–97.
[55] A. Chambolle, T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, J. Math. Imag. Vision 40 (1) (2011) 120–145.
[56] T.F. Chan, S.H. Kang, J. Shen, Euler's elastica and curvature-based inpainting, SIAM J. Appl. Math. (2002) 564–592.
[57] H. Chao, Y. Gu, M. Napolitano, A survey of optical flow techniques for robotics navigation applications, J. Intell. Robotic Syst. 73 (1–4) (2014) 361–372.
[58] P. Charbonnier, L. Blanc-Féraud, G. Aubert, M. Barlaud, Deterministic edge-preserving regularization in computed imaging, IEEE Trans. Image Process. 6 (2) (1997) 298–311.
[59] Z. Chen, H. Jin, Z. Lin, S. Cohen, Y. Wu, Large displacement optical flow from nearest neighbor fields, in: Computer Vision and Pattern Recognition (CVPR), 2013, pp. 2443–2450.
[60] T.M. Chin, W.C. Karl, A.S. Willsky, Probabilistic and sequential computation of optical flow using temporal coherence, IEEE Trans. Image Process. 3 (6) (1994) 773–788.
[61] P. Chou, C. Brown, The theory and practice of Bayesian image labeling, Int. J. Comput. Vision 4 (3) (1990) 185–210.
[62] P.G. Ciarlet, The Finite Element Method for Elliptic Problems, Elsevier, 1978.
[63] I. Cohen, Nonlinear variational method for optical flow computation, in: Proceedings of the 8th Scandinavian Conference on Image Analysis, 1993, pp. 523–530.
[64] T. Cooke, Two applications of graph-cuts to image processing, in: Digital Image Computing: Techniques and Applications (DICTA), 2008, pp. 498–504.
[65] T. Corpetti, É. Mémin, P. Pérez, Dense estimation of fluid flows, IEEE Trans. Pattern Anal. Mach. Intell. 24 (3) (2002) 365–380.
[66] D. Cremers, S. Soatto, Motion competition: a variational approach to piecewise parametric motion segmentation, Int. J. Comput. Vision 62 (3) (2005) 249–265.
[67] A. Crétual, F. Chaumette, Visual servoing based on image motion, Int. J. Robotics Res. 20 (11) (2001) 857–877.
[68] A. Criminisi, P. Pérez, K. Toyama, Region filling and object removal by exemplar-based image inpainting, IEEE Trans. Image Process. 13 (9) (2004) 1200–1212.
[69] T. Crivelli, B. Cernuschi-Frias, P. Bouthemy, J.-F. Yao, Motion textures: modeling, classification, and segmentation using mixed-state Markov random fields, SIAM J. Imag. Sci. 6 (4) (2013) 2484–2520.
[70] R. Cutler, M. Turk, View-based interpretation of real-time optical flow for gesture recognition, in: International Conference on Automatic Face and Gesture Recognition, 1998, pp. 416–416.
[71] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Computer Vision and Pattern Recognition (CVPR), vol. 1, 2005, pp. 886–893.
[72] J. Delon, B. Rougé, Small baseline stereovision, J. Math. Imag. Vision 28 (3) (2007) 209–223.
[73] J. Delpiano, J. Jara, J. Scheer, O.A. Ramírez, J. Ruiz-del Solar, S. Härtel, Performance of optical flow techniques for motion analysis of fluorescent point signals in confocal microscopy, Mach. Vis. Appl. 23 (4) (2012) 675–689.
[74] O. Demetz, M. Stoll, S. Volz, J. Weickert, A. Bruhn, Learning brightness transfer functions for the joint recovery of illumination changes and optical flow, in: European Conference on Computer Vision (ECCV), Zürich, September 2014, pp. 455–471.
[75] P. Dérian, P. Héas, C. Herzet, É. Mémin, Wavelet-based fluid motion estimation, in: Scale Space and Variational Methods in Computer Vision (SSVM), 2011, pp. 737–748.
[76] P. Dérian, P. Héas, C. Herzet, E. Memin, et al., Wavelets and optical flow motion estimation, Numer. Math.: Theory Methods Appl. 6 (1) (2013) 117–137.
[77] R. Deriche, P. Kornprobst, G. Aubert, Optical-flow estimation while preserving its discontinuities: a variational approach, in: Recent Developments in Computer Vision, LNCS, vol. 1035, Springer, 1996, pp. 69–80.
[78] M. Drulea, S. Nedevschi, Total variation regularization of local-global optical flow, in: Intelligent Transportation Systems Conference (ITSC), 2011, pp. 318–323.
[79] M. Drulea, S. Nedevschi, Motion estimation using the correlation transform, IEEE Trans. Image Process. 22 (8) (2013) 3260–3270.
[80] R. Dupont, N. Paragios, R. Keriven, P. Fuchs, Extraction of layers of similar motion through combinatorial techniques, in: Energy Minimization Methods in Computer Vision and Pattern Recognition, Springer, 2005, pp. 220–234.
[81] W. Enkelmann, Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences, Comput. Vision Graph. Image Process. 43 (2) (1988) 150–177.
[82] W. Enkelmann, Obstacle detection by evaluation of optical flow fields from image sequences, Image Vis. Comput. 9 (3) (1991) 160–168.
[83] G. Facciolo, N. Limare, E. Meinhardt, Integral images for block matching, Image Process. On Line (2013).
[84] S. Fazekas, T. Amiaz, D. Chetverikov, N. Kiryati, Dynamic texture detection based on motion analysis, Int. J. Comput. Vision 82 (1) (2009) 48–63.
[85] S. Fazekas, D. Chetverikov, Normal versus complete flow in dynamic texture recognition: a comparative study, in: Workshop on Texture Analysis and Synthesis at International Conference on Computer Vision (ICCV), Beijin, China, 2005, pp. 37–42.
[86] A. Fix, A. Gruber, E. Boros, R. Zabih, A graph cut algorithm for higher-order Markov random fields, in: International Conference on Computer Vision (ICCV), 2011, pp. 1020–1027.
[87] D.J. Fleet, M.J. Black, Y. Yacoob, A.D. Jepson, Design and use of linear models for image motion analysis, Int. J. Comput. Vision 36 (3) (2000) 171–193.
[88] D. Fortun, P. Bouthemy, C. Kervrann, Aggregation of local parametric candidates with exemplar-based occlusion handling for optical flow, preprint arXiv:1407.5759.
[89] D. Fortun, P. Bouthemy, P. Paul-Gilloteaux, C. Kervrann, Aggregation of patch-based estimations for illumination-invariant optical flow in live cell imaging, in: International Symposium on Biomedical Imaging (ISBI), IEEE, 2013, pp. 660–663.
[90] D. Fortun, C. Chen, P. Paul-Gilloteaux, F. Waharte, J. Salamero, C. Kervrann, Correlation and variational approaches for motion and diffusion estimation in fluorescence imaging, in: European Signal Processing Conference (EUSIPCO), September 2013, pp. 660–663.
[91] D.R. Fulkerson, Flows in Networks, Princeton University Press, 1962.
[92] R. Gal, N. Kiryati, N. Sochen, Progress in the restoration of image sequences degraded by atmospheric turbulence, Pattern Recogn. Lett. 14 (2014) 8–14.
[93] R. Garg, A. Roussos, L. Agapito, Robust trajectory-space TV-L1 optical flow for non-rigid sequences, in: Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), Springer, 2011, pp. 300–314.
[94] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving? The KITTI vision benchmark suite, in: Computer Vision and Pattern Recognition (CVPR), 2012, pp. 3354–3361.
[95] M. Gelgon, P. Bouthemy, T. Dubois, A region tracking method with failure detection for an interactive video indexing environment, in: Visual Information and Information Systems, Springer, 1999, pp. 261–269.
[96] D. Geman, G. Reynolds, Constrained restoration and the recovery of discontinuities, IEEE Trans. Pattern Anal. Mach. Intell. 14 (3) (1992) 367–383.
[97] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Trans. Pattern Anal. Mach. Intell. (6) (1984) 721–741.
[98] A. Giachetti, M. Campani, V. Torre, The use of optical flow for road navigation, IEEE Trans. Robotics Autom. 14 (1) (1998) 34–48.
[99] B. Glocker, T. Heibel, N. Navab, P. Kohli, C. Rother, Triangleflow: optical flow with triangulation-based higher-order likelihoods, in: European Conference on Computer Vision (ECCV), 2010, pp. 272–285.
[100] B. Glocker, N. Komodakis, N. Paragios, G. Tziritas, N. Navab, Inter and intra-modal deformable registration: continuous deformations meet efficient optimal linear programming, in: Information Processing in Medical Imaging (IPMI), Springer, 2007, pp. 408–420.
[101] B. Glocker, N. Paragios, N. Komodakis, G. Tziritas, N. Navab, Optical flow estimation with uncertainties through dynamic MRFs, in: Computer Vision and Pattern Recognition (CVPR), Anchorage, Alaska, June 2008, pp. 1–8.
[102] A.V. Goldberg, R.E. Tarjan, A new approach to the maximum-flow problem, J. ACM (JACM) 35 (4) (1988) 921–940.
[103] P. Golland, A.M. Bruckstein, Motion from color, Comput. Vis. Image Underst. 68 (3) (1997) 346–362.
[104] S. Grauer-Gray, C. Kambhamettu, Hierarchical belief propagation to reduce search space using cuda for stereo and motion estimation, in: Workshop on Applications of Computer Vision (WACV), 2009, pp. 1–8.
[105] S. Grewenig, J. Weickert, A. Bruhn, From box filtering to fast explicit diffusion, in: DAGM Symposium on Pattern Recognition, 2010, pp. 533–542.
[106] M.D. Grossberg, S.K. Nayar, Modeling the space of camera response functions, IEEE Trans. Pattern Anal. Mach. Intell. 26 (10) (2004) 1272–1282.

[107] P. Gwosdek, H. Zimmer, S. Grewenig, A. Bruhn, J. Weickert, A highly efficient GPU implementation for variational optic flow based on the Euler–Lagrange framework, in: Trends and Topics in Computer Vision, 2012, pp. 372–383.

[108] D. Hafner, O. Demetz, J. Weickert, Why is the census transform good for robust optic flow computation?, in: Scale Space and Variational Methods in Computer Vision (SSVM), 2013, pp. 210–221.

[109] N. Hata, A. Nabavi, W.M. Wells III, S.K. Warfield, R. Kikinis, P.M. Black, F.A. Jolesz, Three-dimensional optical flow method for measurement of volumetric brain deformation from intraoperative mr images, J. Comput. Assist. Tomogr. 24 (4) (2000) 531–538.

[110] H.W. Haussecker, D.J. Fleet, Computing optical flow with physical models of brightness variation, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 661–673.

[111] K. He, J. Sun, X. Tang, Guided image filtering, in: European Conference on Computer Vision (ECCV), 2010, pp. 1–14.

[112] P. Héas, C. Herzet, E. Memin, Bayesian inference of models and hyperparameters for robust optical-flow estimation, IEEE Trans. Image Process. 21 (4) (2012) 1437–1451.

[113] P. Héas, E. Mémin, N. Papadakis, A. Szantai, Layered estimation of atmospheric mesoscale dynamics from satellite imagery, Trans. Geosci. Remote Sens. 45 (12) (2007) 4087–4104.

[114] P. Heise, S. Klose, B. Jensen, A. Knoll, PM-Huber: patchmatch with Huber regularization for stereo matching, in: International Conference on Computer Vision (ICCV), 2013.

[115] D. Heitz, E. Mémin, C. Schnörr, Variational fluid flow measurements from image sequences: synopsis and perspectives, Experim. Fluids 48 (3) (2010) 369–393.

[116] F. Heitz, P. Bouthemy, Multimodal estimation of discontinuous optical flow using Markov random fields, IEEE Trans. Pattern Anal. Mach. Intell. 15 (12) (1993) 1217–1232.

[117] F. Heitz, P. P érez, P. Bouthemy, Multiscale minimization of global energy functions in some visual recovery problems, CVGIP: Image Understand. 59 (1) (1994) 125–134.

[118] S. Hermann, R. Klette, Hierarchical scan-line dynamic programming for optical flow using semi-global matching, in: Workshops of Asian Conference Computer Vision, Springer, 2013, pp. 556–567.

[119] P.W. Holland, R.E. Welsch, Robust regression using iteratively reweighted least-squares, Commun. Stat.-Theory Methods 6 (9) (1977) 813–827.

[120] B. Horn, B. Schunck, Determining optical flow, Artif. Intell. 17 (1–3) (1981) 185–203.

[121] M. Hornácek, F. Besse, J. Kautz, A. Fitzgibbon, C. Rother, Highly overparameterized optical flow using PatchMatch belief propagation, in: European Conference on Computer Vision (ECCV), LNCS, Part III, Zürich, vol. 8689, September 2014, pp. 220–234.

[122] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, M. Gelautz, Fast cost-volume filtering for visual correspondence and beyond, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2) (2013) 504–511.

[123] W. Hu, N. Xie, L. Li, X. Zeng, S. Maybank, A survey on visual content-based video indexing and retrieval, IEEE Trans. Syst. Man Cybern. 41 (6) (2011) 797–819.

[124] T.-C. Huang, C.-K. Chang, C.-H. Liao, Y.-J. Ho, Quantification of blood flow in internal cerebral artery by optical flow method on digital subtraction angiography in comparison with time-of-flight magnetic resonance angiography, PloS one 8 (1) (2013) 54678.

[125] P.J. Huber, Robust Statistics, Springer, 1981.

[126] A. Humayun, O. Mac Aodha, G.J. Brostow, Learning to find occlusion regions, in: Computer Vision and Pattern Recognition (CVPR), 2011, pp. 2161–2168.

[127] S. Ince, J. Konrad, Occlusion-aware optical flow estimation, IEEE Trans. Image Process. 17 (8) (2008) 1443–1451.

[128] H. Ishikawa, Higher-order clique reduction in binary graph cut, in: Computer Vision and Pattern Recognition (CVPR), 2009, pp. 2993–3000.

[129] N. Jacobson, Y. Freund, T. Nguyen, An online learning approach to occlusion boundary detection, IEEE Trans. Image Process. 21 (1) (2012) 252–261.

[130] M. Jain, H. Jégou, P. Bouthemy, Better exploiting motion for better action recognition, in: Computer Vision and Pattern Recognition (CVPR), Portland, Oregon, June 2013, pp. 2555–2562.

[131] M. Jakubowski, G. Pastuszak, Block-based motion estimation algorithms a survey, Opto-Electron. Rev. 21 (1) (2013) 86–102.

[132] P.-M. Jodoin, M. Mignotte, Optical-flow based on an edge-avoidance procedure, Comput. Vis. Image Underst. 113 (4) (2009) 511–531.

[133] S. Ju, M. Black, A. Jepson, Skin and bones: multi-layer, locally affine, optical flow and regularization with transparency, in: Computer Vision and Pattern Recognition (CVPR), 1996, pp. 307–314.

[134] E.M. Kalmoun, L. Garrido, Trust region versus line search for computing the optical flow, Multiscale Model. Simul. 11 (3) (2013) 890–906.

[135] E.M. Kalmoun, L. Garrido, V. Caselles, Line search multilevel optimization as computational methods for dense optical flow, SIAM J. Imag. Sci. 4 (2) (2011) 695–722.

[136] J.H. Kappes, B. Andres, F.A. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B.X. Kausler, J. Lellmann, N. Komodakis, et al., A comparative study of modern inference techniques for discrete energy minimization problems, in: Computer Vision and Pattern Recognition (CVPR), 2013.

[137] C. Kervrann, J. Boulanger, T. Pécot, P. Pérez, J. Salamero, Multiscale neighborhood-wise decision fusion for redundancy detection in image pairs, SIAM J. Multiscale Model. Simul. 9 (4) (2011) 1829–1865.

[138] C. Kervrann, F. Heitz, Statistical deformable model-based segmentation of image motion, Trans. Image Process. 8 (4) (1999) 583–588.

[139] I.-H. Kim, Y.-C.M. Chen, D.L. Spector, R. Eils, K. Rohr, Nonrigid registration of 2-d and 3-d dynamic cell nuclei images for improved classification of subcellular particle motion, IEEE Trans. Image Process. 20 (4) (2011) 1011–1022.

[140] T.H. Kim, H.S. Lee, K.M. Lee, Optical flow via locally adaptive fusion of complementary data costs, in: International Conference on Computer Vision (ICCV), 2013.

[141] Y. Kim, A. Martínez, A. Kak, A local approach for robust optical flow estimation under varying illumination, in: British Machine Vision Conference (BMVC), 2004, pp. 1–10.

[142] Y.-H. Kim, A.M. Martínez, A.C. Kak, Robust motion estimation under varying illumination, Image Vis. Comput. 23 (4) (2005) 365–375.

[143] N. Kiryati, T.R. Raviv, Y. Ivanchenko, S. Rochel, Real-time abnormal motion detection in surveillance video, in: International Conference on Pattern Recognition (ICPR), 2008, pp. 1–4.

[144] P. Kohli, M.P. Kumar, P.H. Torr, $P^3$ & beyond: move making algorithms for solving higher order functions, IEEE Trans. Pattern Anal. Mach. Intell. 31 (9) (2009) 1645–1656.

[145] P. Kohli, P.H. Torr, Dynamic graph cuts for efficient inference in Markov random fields, IEEE Trans. Pattern Anal. Mach. Intell. 29 (12) (2007) 2079–2088.

[146] D.L. Kolin, P.W. Wiseman, Advances in image correlation spectroscopy: measuring number densities, aggregation states, and dynamics of fluorescently labeled macromolecules in cells, Cell Biochem. Biophys. 49 (3) (2007) 141–164.

[147] V. Kolmogorov, Convergent tree-reweighted message passing for energy minimization, IEEE Trans. Pattern Anal. Mach. Intell. 28 (10) (2006) 1568–1583.

[148] V. Kolmogorov, C. Rother, Minimizing nonsubmodular functions with graph cuts-a review, IEEE Trans. Pattern Anal. Mach. Intell. 29 (7) (2007) 1274–1279.

[149] V. Kolmogorov, R. Zabih, Computing visual correspondences with occlusion using graph cuts, in: IEEE Int. Conference on Computer Vision (ICCV), July 2001.

[150] N. Komodakis, N. Paragios, Beyond pairwise energies: efficient optimization for higher-order MRFs, in: Computer Vision and Pattern Recognition (CVPR), 2009, pp. 2985–2992.

[151] N. Komodakis, G. Tziritas, Image completion using efficient belief propagation via priority scheduling and dynamic pruning, IEEE Trans. Image Process. 16 (11) (2007) 2649–2661.

[152] N. Komodakis, G. Tziritas, N. Paragios, Performance vs computational efficiency for optimizing single and dynamic MRFs: setting the state of the art with primal-dual strategies, Comput. Vis. Image Underst. 112 (1) (2008) 14–29.

[153] J. Konrad, E. Dubois, Bayesian estimation of motion vector fields, IEEE Trans. Pattern Anal. Mach. Intell. 14 (9) (1992) 910–927.

[154] P. Krähenbühl, V. Koltun, Efficient nonlocal regularization for optical flow, in: European Conference on Computer Vision (ECCV), 2012, pp. 356–369.

[155] K. Krajsek, R. Mester, On the equivalence of variational and statistical differential motion estimation, in: Southwest Symposium on Image Analysis and Interpretation, 1996, pp. 11–15.

[156] K. Krajsek, R. Mester, Bayesian model selection for optical flow estimation, Pattern Recogn. 4713 (2007) 142–151.

[157] F.R. Kschischang, B.J. Frey, H.-A. Loeliger, Factor graphs and the sum-product algorithm, IEEE Trans. Inform. Theory 47 (2) (2001) 498–519.

[158] N. Kumar, L. Zhang, S. Nayar, What is a good nearest neighbors algorithm for finding similar patches in images?, in: European Conference on Computer Vision (ECCV), 2008, pp. 364–378.

[159] S.-H. Lai, Robust image matching under partial occlusion and spatially varying illumination change, Comput. Vis. Image Underst. 78 (1) (2000) 84–98.

[160] K. Lee, D. Kwon, I. Yun, S. Lee, Optical flow estimation with adaptive convolution kernel prior on discrete framework, in: Computer Vision and Pattern Recognition (CVPR), San Fransisco, June 2010, pp. 2504–2511.

[161] V. Lempitsky, C. Rother, S. Roth, A. Blake, Fusion moves for Markov random field optimization, IEEE Trans. Pattern Anal. Mach. Intell. 32 (8) (2010) 1392–1405.

[162] M. Leordeanu, A. Zanfir, C. Sminchisescu, Locally affine sparse-to-dense matching for motion and occlusion estimation, in: International Conference on Computer Vision (ICCV), 2013.

[163] J. Lewis, Fast normalized cross-correlation, Vision Interface (1995) 120–123.

[164] W. Li, D. Cosker, M. Brown, R. Tang, Optical flow estimation using Laplacian mesh energy, in: Computer Vision and Pattern Recognition (CVPR), June 2013, pp. 2435–2442.

[165] C. Liu, Beyond Pixels: Exploring New Representations and Applications for Motion Analysis, PhD Thesis, Massachusetts Institute of Technology, 2009.

[166] C. Liu, J. Yuen, A. Torralba, Sift flow: dense correspondence across scenes and its applications, IEEE Trans. Pattern Anal. Mach. Intell. 33 (5) (2011) 978–994.

[167] K. Liu, S.S. Lienkamp, A. Shindo, J.B. Wallingford, G. Walz, O. Ronneberger, Optical flow guided cell segmentation and tracking in developing tissue, in: International Symposium on Biomedical Imaging (ISBI), 2014, pp. 298–301.

[168] T. Liu, L. Shen, Fluid flow and optical flow, J. Fluid Mech. 614 (2008) 253–291.

[169] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision 60 (2) (2004) 91–110.

[170] B. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: International Joint Conference on Artificial Intelligence, 1981, pp. 674–679.
[171] J. Luo, E.E. Konofagou, A fast normalized cross-correlation calculation method for motion estimation, IEE Trans. Ultrason. Ferroelectr. Frequency Control 57 (6) (2010) 1347–1357.
[172] Z. Ma, K. He, Y. Wei, J. Sun, E. Wu, Constant time weighted median filtering for stereo matching and beyond, in: International Conference on Computer Vision (ICCV), 2013.
[173] M. Maurizot, P. Bouthemy, B. Delyon, A. Juditski, J.M. Odobez, Determination of singular points in 2d deformable flow fields, in: International Conference on Image Processing (ICIP), vol. 3, 1995, pp. 488–491.
[174] E. Mémin, P. Pérez, Dense estimation and object-based segmentation of the optical flow with robust techniques, IEEE Trans. Image Process. 7 (5) (1998) 703–719.
[175] E. Mémin, P. Pérez, Hierarchical estimation and segmentation of dense motion fields, Int. J. Comput. Vision 46 (2) (2002) 129–155.
[176] Y. Mileva, A. Bruhn, J. Weickert, Illumination-robust variational optical flow with photometric invariants, Pattern Recogn. (2007) 152–162.
[177] A. Mitiche, J.K. Aggarwal, Computer Vision Analysis of Image Motion by Variational Methods, Springer, 2014.
[178] A. Mitiche, P. Bouthemy, Computation and analysis of image motion: a synopsis of current problems and methods, Int. J. Comput. Vision 19 (1) (1996) 29–55.
[179] M.A. Mohamed, B. Mertsching, TV-L1 optical flow estimation with image details recovering based on modified census transform, in: Advances in Visual Computing, 2012, pp. 482–491.
[180] J. Molnár, D. Chetverikov, S. Fazekas, Illumination-robust variational optical flow using cross-correlation, Comput. Vis. Image Underst. 114 (10) (2010) 1104–1114.
[181] M. Mozerov, Constrained optical flow estimation as a matching problem, IEEE Trans. Image Process. 22 (5) (2013) 2044–2055.
[182] T. Müller, C. Rabe, J. Rannacher, U. Franke, R. Mester, Illumination-robust dense optical flow using census signatures, in: Pattern Recogn., LNCS, vol. 6835, Springer, 2011, pp. 236–245.
[183] T. Muller, J. Rannacher, C. Rabe, U. Franke, Feature-and depth-supported modified total variation optical flow for 3d motion field estimation in real scenes, in: Computer Vision and Pattern Recognition (CVPR), 2011, pp. 1193–1200.
[184] D. Mumford, J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems, Commun. Pure Appl. Math. 42 (5) (1989) 577–685.
[185] D.W. Murray, B.F. Buxton, Scene segmentation from visual motion using global optimization, IEEE Trans. Pattern Anal. Mach. Intell. 9 (2) (1987) 220–228.
[186] H. Nagel, W. Enkelmann, An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences, IEEE Trans. Pattern Anal. Mach. Intell. 8 (5) (1986) 565–593.
[187] H.-H. Nagel. Extending the oriented smoothness constraint into the temporal domain and the estimation of derivatives of optical flow, in: European Conference on Computer Vision (ECCV), 1990, pp. 139–148.
[188] S. Negahdaripour, Revised definition of optical flow: integration of radiometric and geometric cues for dynamic scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (9) (1998) 961–979.
[189] L. Ng, V. Solo, Errors-in-variables modelling in optical flow problems, in: International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 5, 1998, pp. 2773–2776.
[190] C. Nieuwenhuis, D. Kondermann, C.S. Garbe, Complex motion models for simple optical flow estimation, in: Pattern Recogn., LNCS, vol. 6376, Springer, 2010, pp. 141–150.
[191] T. Nir, A.M. Bruckstein, R. Kimmel, Over-parameterized variational optical flow, Int. J. Comput. Vision 76 (2) (2008) 205–216.
[192] P. Ochs, Y. Chen, T. Brox, T. Pock, iPiano: inertial proximal algorithm for non-convex optimization, SIAM J. Imag. Sci. 7 (2) (2013) 1388–1419.
[193] J.M. Odobez, P. Bouthemy, Robust multiresolution estimation of parametric motion models, J. Visual Commun. Image Represent. 6 (4) (1995) 348–365.
[194] J.M. Odobez, P. Bouthem, Separation of moving regions from background in an image sequence acquired with a mobile camera, in: H.-H. Li, S. Sun, H. Derin (Eds.), Video Data Compression for Multimedia Computing, Series in Engineering and Computer Science, vol. 378, Springer, 1997, pp. 283–311 (Chapter 8).
[195] J.M. Odobez, P. Bouthemy, Direct incremental model-based image motion segmentation for video analysis, Signal Process. 66 (2) (1998) 143–155.
[196] G. Panin, Mutual information for multi-modal, discontinuity-preserving image registration, in: Advances in Visual Computing, 2012, pp. 70–81.
[197] N. Papadakis, R. Yildizoglu, J.-F. Aujol, V. Caselles, High-dimension multilabel problems: convex or nonconvex relaxation?, SIAM J Imag. Sci. 6 (4) (2013) 2603–2639.
[198] N. Papenberg, A. Bruhn, T. Brox, S. Didas, J. Weickert, Highly accurate optic flow computation with theoretically justified warping, Int. J. Comput. Vision 67 (2) (2006) 141–158.
[199] N. Paragios, R. Deriche, Geodesic active regions and level set methods for motion estimation and tracking, Comput. Vis. Image Underst. 97 (3) (2005) 259–282.
[200] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufman, 1988.
[201] M. Pingault, E. Bruno, D. Pellerin, A robust multisacle B-spline function decomposition for estimating motion transparency, IEEE Trans. Image Process. 12 (11) (2003) 1416–1426.
[202] G. Piriou, P. Bouthemy, J.-F. Yao, Recognition of dynamic video contents with global probabilistic models of visual motion, IEEE Trans. Image Process. 15 (11) (2006) 3417–3430.
[203] L. Pizarro, J. Delpiano, P. Aljabar, J. Ruiz-del Solar, D. Rueckert, Towards dense motion estimation in light and electron microscopy, in: International Symposium on Biological Imaging (ISBI), 2011, pp. 1939–1942.
[204] T. Pock, M. Pock, H. Bischof, Algorithmic differentiation: application to variational problems in computer vision, IEEE Trans. Pattern Anal. Mach. Intell. 29 (7) (2007) 1180–1193.
[205] R. Ranftl, S. Gehrig, T. Pock, H. Bischof, Pushing the limits of stereo using variational stereo estimation, in: Intelligent Vehicles Symposium (IV), 2012, pp. 401–407.
[206] H.A. Rashwan, M.A. García, D. Puig, Variational optical flow estimation based on stick tensor voting, IEEE Trans. Image Process. 22 (7) (2013) 589–2599.
[207] T. Riklin-Raviv, N. Sochen, N. Kiryati, Shape-based mutual segmentation, Int. J. Comput. Vision 79 (3) (2008) 231–245.
[208] S. Roth, M. Black, On the spatial statistics of optical flow, Int. J. Comput. Vision 74 (1) (2007) 33–50.
[209] C. Rother, V. Kolmogorov, V. Lempitsky, M. Szummer, Optimizing binary MRFs via extended roof duality, in: Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1–8.
[210] L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D 60 (1) (1992) 259–268.
[211] D. Rueckert, L.I. Sonoda, C. Hayes, D.L. Hill, M.O. Leach, D.J. Hawkes, Nonrigid registration using free-form deformations: application to breast MR images, IEEE Trans. Pattern Anal. Mach. Intell. 18 (8) (1999) 712–721.
[212] J.A. Schnabel, D. Rueckert, M. Quist, J.M. Blackall, A.D. Castellano-Smith, T. Hartkens, G.P. Penney, W.A. Hall, H. Liu, C.L. Truwit, et al., A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, 2001, pp. 573–581.
[213] C. Schnörr, W. Peckar, Motion-based identification of deformable templates, in: International Conference Computer Analysis of Images and Patterns (CAIP), Prague, 1995, pp. 122–129.
[214] T. Schoenemann, D. Cremers, Near real-time motion segmentation using graph cuts, in: DAGM Symposium on Pattern Recognition, 2006, pp. 455–464.
[215] T. Senst, V. Eiselen, T. Sikora, Robust local optical flow for feature tracking, IEEE Trans. Circ. Syst. Video Technol. 22 (9) (2012) 1377–1387.
[216] A. Shekhovtsov, I. Kovtun, V. Hlaváč, Efficient MRF deformation model for non-rigid image matching, Comput. Vis. Image Underst. 112 (1) (2008) 91–99.
[217] X. Shen, Y. Wu, Sparsity model for robust optical flow estimation at motion discontinuities, in: Computer Vision and Pattern Recognition (CVPR), 2010, pp. 2456–2463.
[218] J. Shi, C. Tomasi, Good features to track, in: Computer Vision and Pattern Recognition (CVPR), 1994, pp. 593–600.
[219] W. Shi, X. Zhuang, L. Pizarro, W. Bai, H. Wang, K.-P. Tung, P. Edwards, D. Rueckert, Registration using sparse free-form deformations, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2012, pp. 659–666.
[220] D. Shulman, J.-Y. Herve, Regularization of discontinuous flow fields, in: Workshop on Visual Motion, 1989, pp. 81–86.
[221] E.P. Simoncelli, E.H. Adelson, D.J. Heeger, Probability distributions of optical flow, in: Computer Vision and Pattern Recognition (CVPR), 1991, pp. 310–315.
[222] S.N. Sinha, J.-M. Frahm, M. Pollefeys, Y. Genc, Feature tracking and matching in video using programmable graphics hardware, Mach. Vis. Appl. 22 (1) (2011) 207–217.
[223] P. Smith, T. Drummond, R. Cipolla, Layered motion segmentation and depth ordering by tracking edges, IEEE Trans. Pattern Anal. Mach. Intell. 26 (4) (2004) 479–494.
[224] A. Sotiras, C. Davatzikos, N. Paragios, Deformable medical image registration: a survey, IEEE Trans. Med. Imag. 32 (7) (2013) 1153–1190.
[225] A.N. Stein, M. Hebert, Occlusion boundaries from motion: low-level detection and mid-level reasoning, Int. J. Comput. Vision 82 (3) (2009) 325–357.
[226] F. Stein, Efficient computation of optical flow using the census transform, in: Pattern Recognition, LNCS, vol. 3175, Springer, 2004, pp. 79–86.
[227] F. Steinbrucker, T. Pock, D. Cremers, Advanced data terms for variational optic flow estimation, in: Vision, Modeling, and Visualization Workshop, 2009.
[228] C. Stiller, J. Konrad, Estimating motion in image sequences, IEEE Signal Process. Mag. 16 (4) (1999) 70–91.
[229] C. Strecha, R. Fransens, L. Van Gool, A probabilistic approach to large displacement optical flow and occlusion detection, in: Statistical Methods in Video Processing, Springer, 2004, pp. 71–82.
[230] C.-W. Su, H.-Y. Liao, H.-R. Tyan, C.-W. Lin, D.-Y. Chen, K.-C. Fan, Motion flow-based video retrieval, IEEE Trans. Multimedia 9 (6) (2007) 1193–1201.
[231] D. Sun, S. Roth, M. Black, Secrets of optical flow estimation and their principles, in: Computer Vision and Pattern Recognition (CVPR), San Fransisco, June 2010, pp. 2432–2439.
[232] D. Sun, S. Roth, J. Lewis, M. Black, Learning optical flow, in: European Conference on Computer Vision (ECCV), 2008, pp. 83–97.
[233] D. Sun, E. Sudderth, M. Black, Layered image motion with explicit occlusions, temporal consistency, and depth ordering, in: Advances in Neural

Information Processing Systems (NIPS), Vancouver, December 2010, pp. 2226–2234.
[234] D. Sun, E.B. Sudderth, M.J. Black, Layered segmentation and optical flow estimation over time, in: Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1768–1775.
[235] Z. Sun, G. Bebis, R. Miller, On-road vehicle detection: a review, IEEE Trans. Pattern Anal. Mach. Intell. 28 (5) (2006) 694–711.
[236] R. Szeliski, H.-Y. Shum, Motion estimation with quadtree splines, IEEE Trans. Pattern Anal. Mach. Intell. 18 (12) (1996) 1199–1210.
[237] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, C. Rother, A comparative study of energy minimization methods for markov random fields with smoothness-based priors, IEEE Trans. Pattern Anal. Mach. Intell. 30 (6) (2008) 1068–1080.
[238] M. Tao, J. Bai, P. Kohli, S. Paris, Simpleflow: a non-iterative, sublinear optical flow algorithm, in: Computer Graphics Forum, vol. 31, 2012, pp. 345–353.
[239] C.-H. Teng, S.-H. Lai, Y.-S. Chen, W.-H. Hsu, Accurate optical flow computation under non-uniform brightness variations, Comput. Vis. Image Underst. 97 (3) (2005) 315–346.
[240] M. Tistarelli, Multiple constraints to compute optical flow, IEEE Trans. Pattern Anal. Mach. Intell. 18 (12) (1996) 1243–1250.
[241] Trobin, T. Pock, D. Cremers, H. Bischof, An unbiased second-order prior for high-accuracy motion estimation, in: DAGM Symposium on Pattern Recognition, 2008, pp. 396–405.
[242] W. Trobin, T. Pock, D. Cremers, H. Bischof, Continuous energy minimization via repeated binary fusion, in: European Conference on Computer Vision (ECCV), Marseille, France, October 2008, pp. 677–690.
[243] D. Tzovaras, M.G. Strintzis, H. Sahinoglou, Evaluation of multiresolution block matching techniques for motion and disparity estimation, Signal Process.: Image Commun. 6 (1) (1994) 59–67.
[244] M. Unger, M. Werlberger, T. Pock, H. Bischof, Joint motion estimation and segmentation of complex scenes with label costs and occlusion modeling, in: Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1878–1885.
[245] S. Uras, F. Girosi, A. Verri, V. Torre, A computational approach to motion perception, Biol. Cybern. 60 (2) (1988) 79–87.
[246] J. van de Weijer, T. Gevers, Robust optical flow from photometric invariants, in: International Conference on Image Processing (ICIP), vol. 3, 2004, pp. 1835–1838.
[247] C. Vásquez, A. Mitiche, R. Laganière, Joint multiresolution segmentation and parametric estimation of image motion by basis function representation and level set evolution, IEEE Trans. Pattern Anal. Mach. Intell. 28 (5) (2006) 782–793.
[248] S. Vedula, S. Baker, P. Rander, R. Collins, T. Kanade, Three-dimensional scene flow, IEEE Int. Conference on Computer Vision (ICCV), vol. 2, IEEE, 1999, pp. 722–729.
[249] O. Veksler, Efficient Graph-based Energy Minimization Methods in Computer Vision, PhD Thesis, Cornell University, 1999.
[250] O. Veksler, Graph cut based optimization for MRFs with truncated convex priors, in: Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1–8.
[251] A. Verri, T. Poggio, Motion field and optical flow: qualitative properties, IEEE Trans. Pattern Anal. Mach. Intell. 11 (5) (1989) 490–498.
[252] C. Vogel, S. Roth, K. Schindler, An evaluation of data costs for optical flow, in: DAGM Symposium on Pattern Recognition, 2013, pp. 343–353.
[253] S. Volz, A. Bruhn, L. Valgaerts, H. Zimmer, Modeling temporal coherence for optical flow, in: International Conference on Computer Vision (ICCV), 2011, pp. 1116–1123.
[254] M.J. Wainwright, T.S. Jaakkola, A.S. Willsky, MAP estimation via agreement on trees: message-passing and linear programming, IEEE Trans. Inform. Theory 51 (11) (2005) 3697–3717.
[255] C. Wang, N. Komodakis, N. Paragios, Markov random field modeling, inference & learning in computer vision & image understanding: a survey, Comput. Vis. Image Underst. 117 (11) (2013) 1610–1627.
[256] H. Wang, A. Kläser, C. Schmid, C.-L. Liu, Dense trajectories and motion boundary descriptors for action recognition, Int. J. Comput. Vision 103 (1) (2013) 60–79.
[257] J. Wang, E. Adelson, Representing moving images with layers, IEEE Trans. Image Process. 3 (5) (1994) 625–638.
[258] Z. Wang, F. Wu, Z. Hu, MSLD: a robust descriptor for line matching, Pattern Recogn. 42 (5) (2009) 941–953.
[259] J. Weber, J. Malik, Robust computation of optical flow in a multi-scale differential framework, Int. J. Comput. Vision 14 (1) (1995) 67–81.
[260] A. Wedel, D. Cremers, Optical flow estimation, in: Stereo Scene Flow for 3D Motion Analysis, Springer, 2011, pp. 5–30 (Chapter 2).
[261] A. Wedel, D. Cremers, T. Pock, H. Bischof, Structure-and motion-adaptive regularization for high accuracy optic flow, in: International Conference on Computer Vision (ICCV), Kyoto, Japan, October 2009, pp. 1663–1668.
[262] A. Wedel, T. Pock, J. Braun, U. Franke, D. Cremers, Duality TV-L1 flow with fundamental matrix prior, in: Image and Vision Computing New-Zealand, 2008, pp. 1–6.
[263] A. Wedel, T. Pock, C. Zach, H. Bischof, D. Cremers, An improved algorithm for TV-L1 optical flow, in: Statistical and Geometrical Approaches to Visual Motion Analysis, 2009, pp. 23–45.
[264] J. Weickert, A. Bruhn, T. Brox, N. Papenberg, A survey on variational optic flow methods for small displacements, in: Mathematical Models for Registration and Applications to Medical Imaging, Springer, 2006, pp. 103–136.
[265] J. Weickert, J. Heers, C. Schnörr, K.J. Zuiderveld, O. Scherzer, H. Siegfried Stiehl, Fast parallel algorithms for a broad class of nonlinear variational diffusion approaches, Real-Time Imag. 7 (1) (2001) 31–45.
[266] J. Weickert, C. Schnörr, A theoretical framework for convex regularizers in PDE-based computation of image motion, Int. J. Comput. Vision 45 (3) (2001) 245–264.
[267] J. Weickert, C. Schnörr, Variational optic flow computation with a spatio-temporal smoothness constraint, J. Math. Imag. Vision 14 (3) (2001) 245–255.
[268] P. Weinzaepfel, J. Revaud, Z. Harchaoui, C. Schmid, Deepflow: large displacement optical flow with deep matching, in: International Conference on Computer Vision (ICCV), 2013.
[269] M. Werlberger, T. Pock, H. Bischof, Motion estimation with non-local total variation regularization, in: Computer Vision and Pattern Recognition (CVPR), San Fransisco, June 2010, pp. 2464–2471.
[270] M. Werlberger, T. Pock, M. Unger, H. Bischof, Optical flow guided TV-L1 video interpolation and restoration, in: Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), 2011, pp. 273–286.
[271] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, H. Bischof, Anisotropic Huber-L1 optical flow, in: British Machine Vision Conference (BMVC), 2009.
[272] M. Werlberger, M. Unger, T. Pock, H. Bischof, Efficient minimization of the non-local Potts model, in: Scale Space and Variational Methods in Computer Vision (SSVM), 2012, pp. 314–325.
[273] O. Williams, M. Isard, J. MacCormick, Estimating disparity and occlusions in stereo video sequences, in: Computer Vision and Pattern Recognition (CVPR), June 2005.
[274] Y.-T. Wu, T. Kanade, C.-C. Li, J. Cohn, Image registration using wavelet-based motion model, Int. J. Comput. Vision 38 (2) (2000) 129–152.
[275] M. Xavier, A. Lalande, P.M. Walker, F. Brunotte, L. Legrand, An adapted optical flow algorithm for robust quantification of cardiac wall motion from standard cine-mr examinations, IEEE Trans. Inform. Technol. Biomed. 16 (5) (2012) 859–868.
[276] J. Xiao, H. Cheng, H. Sawhney, C. Rao, M. Isnardi, Bilateral filtering-based optical flow estimation with occlusion detection, in: European Conference on Computer Vision (ECCV), 2006, pp. 211–224.
[277] J. Xiao, M. Shah, Motion layer extraction in the presence of occlusions using graph cuts, IEEE Trans. Pattern Anal. Mach. Intell. 27 (10) (2005) 1644–1659.
[278] L. Xu, J. Chen, J. Jia, A segmentation-based variational model for accurate optical flow estimation, in: European Conference on Computer Vision (ECCV), Marseille, October 2008, pp. 671–684.
[279] L. Xu, J. Jia, Y. Matsushita, Motion detail preserving optical flow estimation, IEEE Trans. Pattern Anal. Mach. Intell. 34 (9) (2012) 1744–1757.
[280] J. Yang, W. Yin, Y. Zhang, Y. Wang, A fast algorithm for edge-preserving variational multichannel image restoration, SIAM J. Imag. Sci. 2 (2) (2009) 569–592.
[281] K.-J. Yoon, I.S. Kweon, Adaptive support-weight approach for correspondence search, IEEE Trans. Pattern Anal. Mach. Intell. 28 (4) (2006) 650–656.
[282] G. Yu, J.-M. Morel, ASIFT: an algorithm for fully affine invariant comparison, SIAM J. Imag. Sci. 2 (2009) 438–469.
[283] R. Zabih, J. Woodfill, Non-parametric local transforms for computing visual correspondence, in: European Conference on Computer Vision (ECCV), 1994, pp. 151–158.
[284] C. Zach, D. Gallup, J.-M. Frahm, Fast gain-adaptive KLT tracking on the GPU, in: Computer Vision and Pattern Recognition Workshops (CVPRW), 2008, pp. 1–7.
[285] C. Zach, T. Pock, H. Bischof, A duality based approach for realtime TV-L1 optical flow, in: DAGM symposium on Pattern Recognition, 2007, pp. 214–223.
[286] H. Zimmer, A. Bruhn, J. Weickert, Optic flow in harmony, Int. J. Comput. Vision 93 (3) (2011) 1–21.
[287] C. Zitnick, N. Jojic, S. Kang, Consistent segmentation for optical flow estimation, in: International Conference on Computer Vision (ICCV), Beijing, October 2005, pp. 1308–1315.