A survey of variational and CNN-based optical flow techniques

Article · March 2019

DOI: 10.1016/j.image.2018.12.002

7 authors, including:

Zhigang Tu
Wuhan University
33 PUBLICATIONS **353** CITATIONS
SEE PROFILE

Dejun Zhang
China University of Geosciences
47 PUBLICATIONS **518** CITATIONS
SEE PROFILE

Remco C. Veltkamp
Utrecht University
317 PUBLICATIONS **8,440** CITATIONS
SEE PROFILE

Signal Processing: Image Communication 72 (2019) 9–24

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

A survey of variational and CNN-based optical flow techniques

Zhigang Tu [a,*], Wei Xie [b,*], Dejun Zhang [c], Ronald Poppe [d], Remco C. Veltkamp [d], Baoxin Li [e], Junsong Yuan [f]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote sensing, Wuhan University, 430079, Wuhan, China
[b] School of Computer, Central China Normal University, LuoyuRoad 152, Wuhan, China
[c] School of Information Engineering, China University of Geosciences, 30074, Wuhan, China
[d] Department of Information and Computing Sciences, Utrecht University, Princetonplein 5, Utrecht, The Netherlands
[e] School of Computing, Informatics, Decision System Engineering, Arizona State University, AZ 85287, USA
[f] Computer Science and Engineering department, State University of New York at Buffalo, NY 14260-2500, USA

ARTICLE INFO

Keywords:
Optical flow
Variational method
CNN-based method
Evaluation measures
Challenges

ABSTRACT

Dense motion estimations obtained from optical flow techniques play a significant role in many image processing and computer vision tasks. Remarkable progress has been made in both theory and its application in practice. In this paper, we provide a systematic review of recent optical flow techniques with a focus on the variational method and approaches based on Convolutional Neural Networks (CNNs). These two categories have led to state-of-the-art performance. We discuss recent modifications and extensions of the original model, and highlight remaining challenges. For the first time, we provide an overview of recent CNN-based optical flow methods and discuss their potential and current limitations.

1. Introduction

Motion provides essential information for a wide variety of visual tasks, and directly affects the subsequent image processing. Consequently, the computation of motion information from image sequences is an important issue in computer vision and image processing. Optical flow, one of the most successful motion estimation methods, has been widely investigated. Since Horn and Schunck (HS) [1] as well as Lucas and Kanade (LK) [2] proposed the differential method to calculate optical flow in 1981, a great deal of extensions and modifications have been proposed. More recently, a new line of optical flow algorithms considers the use of Convolutional Neural Networks (CNNs). The two classes of algorithm have become the predominant ways to calculate optical flow [3–5]. In this survey, we review both and discuss challenges and opportunities.

1.1. Optical flow

The relative movement of a 3D scene observed with a camera leads to changes in the 2D projected images. The estimation of this motion comes down to the computation of a projection of the actual motion onto the image plane [6]. To this end, algorithms have relied on the analysis of brightness variations of pixels in pairs of subsequent images. The 2D displacement field that describes apparent motion of brightness patterns between two successive images is called the optical flow [1]. The optical flow field is often considered as the projected scene flow field, which is the true motion of the objects in the scene as viewed from the image plane [4]. The optical flow field is ideally a dense field of displacement vectors (see Fig. 1), which maps all points of the first image onto their corresponding locations in the second image.

The concept of optical flow was proposed by Gibson [7]. Poggio and Reiehardt [8] presented an approach to compute the motion of each pixel in an image, which can be considered a rough flow method. A first practical optical flow model was established by the classical work of Horn and Schunck (HS) [1]. It is based on the assumption that the brightness of a pixel keeps constant during a short time interval, which is known as the brightness constancy assumption (BCA).

1.2. Applications of optical flow

Although the optical flow field is an approximate projection of the true motion of the scene, it provides valuable information about the spatial arrangement of the viewed objects and the change rate of the arrangement. We discuss several domains where optical flow is used.

Visual Surveillance. Visual surveillance systems are often designed as modular systems of functional modules such as motion detection, depth estimation, object tracking, and object behavioral analysis [9]. Optical flow is effective in separating the foreground and background, and to identify moving objects [10]. This allows for the detection and

* Corresponding authors.
E-mail addresses: tuzhigang@whu.edu.cn (Z. Tu), XW@mail.ccnu.edu.cn (W. Xie), zhangdejun@cug.edu.cn (D. Zhang), R.W.Poppe@uu.nl (R. Poppe), R.C.Veltkamp@uu.nl (R.C. Veltkamp), Baoxin.Li@asu.edu (B. Li), jsyuan@buffalo.edu (J. Yuan).

https://doi.org/10.1016/j.image.2018.12.002
Received 17 March 2018; Received in revised form 27 October 2018; Accepted 3 December 2018
Available online 12 December 2018

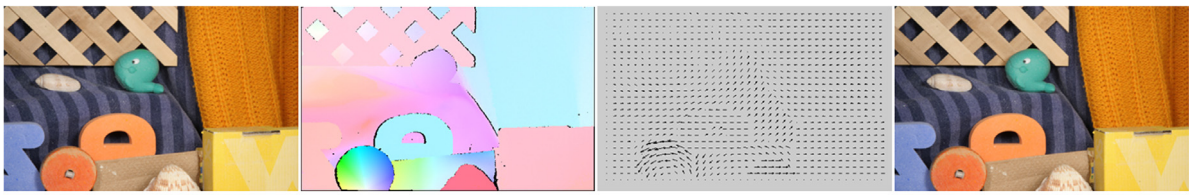

Fig. 1. Illustration of optical flow: (a), (d) frames 10 and 11 of the RubberWhale sequence on the Middlebury benchmark [5], respectively; (b) color-coded ground truth flow (black regions indicate unknown flow vectors); (c) vector plot ground truth flow.

tracking of objects through the scene. Optical flow is extensively used for visual surveillance tasks including tracking [11,12], segmentation [13,14], action recognition [15,16] and anomaly detection [17,18].

Robot Navigation. Navigation can be roughly described as the process of determining a suitable path between a robot's start and goal location [19]. As optical flow is the apparent motion of the brightness patterns of image sequences, it not only provides important information about the unknown environment, but it is also helpful to determine the direction and the speed of the robot. Due to this characteristic, optical flow is one of the two primary techniques for mapless robot navigation [20]. Using optical flow for robot navigation is inspired by emulating the flying behavior of bees [21]. Optical flow is widely used for obstacle detection [22] and collision avoidance [23].

Image Interpolation and Super-Resolution. High definition TVs (HDTVs) have significantly improved our visual experience [24]. Image sequence interpolation [25,26] is the process of creating intermediate images between two given consecutive images, to increase the frame rate. It has broad applications in the fields of video post-processing and restoration. Image sequence super-resolution [27,28] aims at combining the available information from a sequence of low-resolution (LR) images to produce a high-resolution (HR) image with more details. Both image interpolation and super-resolution require accurate and detailed alignment of pixels in which the dense optical flow field plays a crucial role.

Physical and Metrological Applications. Optical flow can be applied as a measurement in modeling and simulation, including weather forecasting [29] and fluid dynamics [30,31].

1.3. Optical flow techniques

In this section, we introduce the primary optical flow techniques and give a detailed analysis of their characteristics.

1.3.1. Differential technique

The differential technique also refers to the gradient-based approach. It computes the velocity from spatial and temporal derivatives of the image brightness [4]. The constraint relies purely on the BCA. This leads to an aperture problem: one equation cannot uniquely determine two unknown components (the horizontal and vertical pixel displacement) of the flow field. An additional constraint encoding *a priori* information on the flow field needs to be incorporated with the BCA to make the problem well-posed [1,2]. Normally, the prior takes the form of spatial coherence imposed by either local or global constraints. Generally, we identify two approaches [32,33]:

- **Global Differential Method** introduces a global smoothness constraint (also called regularization term) to solve the aperture problem. The assumption is that neighboring pixels come from the same object and consequently undergo a similar motion. As a result, the displacement field of the brightness patterns varies smoothly. The variational method proposed by Horn and Schunck [1] is generally considered as the typical global method. It is one of the most successful techniques to compute optical flow [6]. Solving the optical flow estimation requires the minimization of the associated Euler–Lagrange equations. These form a system of partial differential equations that can be derived analytically using the calculus of variations [34]. In practice, the system can be solved with numerical methods such as Successive Over-Relaxation (SOR) [35], or efficient multi-grid methods [36]. In general, there are two main ways to minimize the global energy function. One is a continuous optimization [6,37] such as gradient descent [38,39] and the variational method [1,35]. The second approach is a discrete optimization [40], for example using graph cuts algorithms [41,42] or message-passing algorithms [43,44]. The discrete optimization method approximates the continuous space of solutions and enables a more thorough and complete search of the state space. Since discrete methods do not require differentiation of the energy, they can handle a wider variety of data and regularization terms than continuous methods [3,45]. On the other hand, discrete optimization methods are generally limited in terms of accuracy and efficiency by the number of labels and the size of the label space [46,47]. For optical flow, sub-pixel accuracy is often required [40], and real-time computation is necessary for many applications [48,49].
- **Local Differential Method** assumes that the motion within a local neighborhood can be described by a parametric model [2,50]. For each pixel in a local neighborhood, an equation relating brightness and flow can be derived. The set of equations for all the pixels within the local neighborhood is then solved to estimate the optical flow by performing least-squares minimization. Setting the proper size of the neighborhood is a challenge. A small-sized neighborhood might not contain sufficient information to handle the aperture problem. In contrast, a large-sized neighborhood allows the integration of information over many pixels but may include pixels from other motion surfaces. Moreover, the local method experiences difficulties in homogeneous regions and regions is motion discontinuities [51].

The global and local methods differ in the interaction of neighboring pixels due to the constraints. Global methods constrain a pixel's flow by its neighboring pixels' flow vectors, whereas local methods constrain a pixel's flow by its neighboring pixels' intensity values. Also, the two methods are computated differently. The global method aggregates the constraint residual over the whole field. Therefore, the flow recovery at one pixel relies on the flow recovery at other pixels. In contrast, the local method integrates the flow constraint of one pixel into a linear system, and solves each pixel's flow independently.

1.3.2. Region-based technique

The region-based technique relies on searching matching patches between two consecutive images. When two corresponding patches have the largest correlation, the optical flow is defined as the shift of the patches [4,52]. Popular measures include the sum of absolute differences (SAD), the sum of squared differences (SSD) [1], and cross-correlation metrics [53]. The region-based technique is more robust to noise than differential techniques. In addition, the region-based technique works well even when images are interlaced or decimated.

1.3.3. Feature-based technique

Feature-based techniques attempt to link sparse but discriminative image features in successive images over time [6,54]. This technique ignores ambiguous areas and, as a result, the calculated flow field is sparse but robust. Discriminative features, in particular corners and edges, as well as low-contrast features such as flat regions, can be matched to determine the optical flow [55]. Feature-based methods consists of two steps: feature detection and correspondence matching. Currently, feature-based optical flow methods are widely used, especially for large displacement matching [56]. The technique has two main drawbacks. First, the optical flow is very sparse when the background or the objects contain features that are not discriminative. Second, the selected features may not be reliable and disappear in subsequent frames.

1.3.4. Frequency-based technique

The frequency-based technique, or velocity tuned filters, calculates optical flow using velocity-tuned filters in the Fourier domain. The advantage is that motion-sensitive mechanisms operating on spatiotemporally oriented energy in the Fourier domain can estimate motion that cannot be estimated using matching approaches [57]. For example, the technique can deal with the motion of random dot patterns. According to the output of the velocity-tuned filters, the frequency-based technique can be classified into two groups:

- **Energy-Based Method**. This method is based on the output energy of the filters [4,58]. The energy of a continuous translation in space is concentrated on a plane in the spatiotemporal frequency domain with the orientation related to the velocity [59]. The first computational model for the perception of motion [60] consists of a quadrature Gabor filter pair. Later, the model was extended to compute optical flow [58]. The optical flow is formulated as the least-squares fit of spatiotemporal energy to a plane in frequency space.
- **Phase-Based Method**. It defines the velocity component in the output of band pass velocity tuned Gabor filters [57,61]. The method is based on the decomposition of the original image into band-pass channels, similar to those produced by quadrature-pair filters in steerable pyramids [62]. Multi-scale representations are typically used for flow computation. A further decomposition into orientation bands yields more local constraints with a generally higher Signal-Noise-Ratio (SNR). Phase-based methods can be considered superior to the energy-based method in three aspects. First, subpixel accuracy can be achieved. Second, the velocity resolution is preserved by taking responses from neighboring filters. Finally, the technique is more robust to changes in viewpoint and illumination, as phase information is insensitive to changes in speed and contrast.

1.3.5. CNN-based technique

CNNs have achieved impressive success in a wide variety of image processing tasks, including optical flow estimation. CNNs are increasing used to replace hand-crafted features by learned features [63,64]. The CNN is applied to extract deep features of the input images. These features are then integrated into common optimization algorithms to calculate optical flow [65,66]. End-to-end methods are one class of approach where the CNN is not only used for learning the image features, but also for matching these features in the images [63,67]. In this way, the whole optical flow process is performed by the CNN (see Fig. 2). While CNNS are typically trained in a supervised way by providing pairs of subsequent frames and the resulting flow field, recent work has also addressed the unsupervised training of CNNs for optical flow [64].

1.4. Evaluation measures

Evaluation measures are aimed at revealing properties of optical flow algorithms and helps to better understand the relative strengths and weaknesses of each. They not only provide information to improve the algorithms but also supply insight to select more suitable approaches to handle special challenges such as dealing with noise, large displacements or motion boundaries. In addition, evaluation measures support researchers to establish more realistic and complex benchmarks, and to develop ways to ensure continuous progress. The measurement of the performance of optical flow algorithms remains the subject of scientific debate, and have led to the introduction of various evaluation measures. We discuss the most relevant ones.

1.4.1. Mathematical measures

We define the ground truth (GT) optical flow as $g(x, y) = (u_T, v_T)$ and the estimated flow as $e(x, y) = (u_E, v_E)$. Some mathematical criteria are proposed to evaluate the performance quantitatively.

- Average Angular Error (AAE):

$$AAE = \frac{1}{MN}\sum \arccos\left(\frac{u_T u_E + v_T v_E + 1}{\sqrt{(u_T^2 + v_T^2 + 1)(u_E^2 + v_E^2 + 1)}}\right) \tag{1}$$

  where M and N are the numbers of columns and rows of the image, and MN is the number of pixels. The AAE measure is defined as the flow error deviation of angle between the GT flow and the estimated flow [61]. The drawback of this measure is that errors in small flows are penalized relatively severely.

- Average Endpoint Error (AEE):

$$AEE = \frac{1}{MN}\sum \sqrt{(u_E - u_T)^2 + (v_E - v_T)^2} \tag{2}$$

  This error measure calculates the Euclidean distance between the endpoints of the GT flow and the estimated flow [68,5]. The AEE discounts errors in regions of small flow while strongly penalizing large estimation errors.

- Average Magnitude Error (AME):

$$AME = \frac{1}{MN}\sum |\sqrt{u_E^2 + v_E^2} - \sqrt{u_T^2 + v_T^2}| \tag{3}$$

  The AME is defined as the average magnitude of the velocity difference between the GT flow and the estimated flow [69]. Since it does not account for errors in direction, it is usually used together with the AAE.

- Normalized Average Magnitude Error (NAME):

$$NAME = \frac{1}{MN}\sum \frac{|\sqrt{u_E^2 + v_E^2} - \sqrt{u_T^2 + v_T^2}|}{\sqrt{u_T^2 + v_T^2}} \tag{4}$$

  This measure is defined as the average ratio between the magnitude of the velocity difference and the magnitude of the ground truth. It is somewhat unreliable for small GT vectors [70].

- Error Normal to the Gradient (ENG):

$$ENG_{\perp}((u_E, v_E), (u_T, v_T)) = \|((u_E - u_T), (v_E - v_T))f^{\perp}(x, y)\| \tag{5}$$

  where $f^{\perp}(x, y) = (-\partial_y I(x, y), \partial_x I(x, y))^T$. This error measure aims to examine the effectiveness of an algorithm to compensate for the aperture problem [69]. Recently, it was pointed out that the residual error between the GT and the estimated warped interpolation (I_{warp}), which is interpolated by the estimated optical flow, is also a good way to evaluate the performance [5].

- Root-Mean-Square (RMS) error:

$$RMS = [\frac{1}{MN}\sum_{(x,y)}(I_{warp}(x, y) - I(x, y))^2]^{1/2} \tag{6}$$

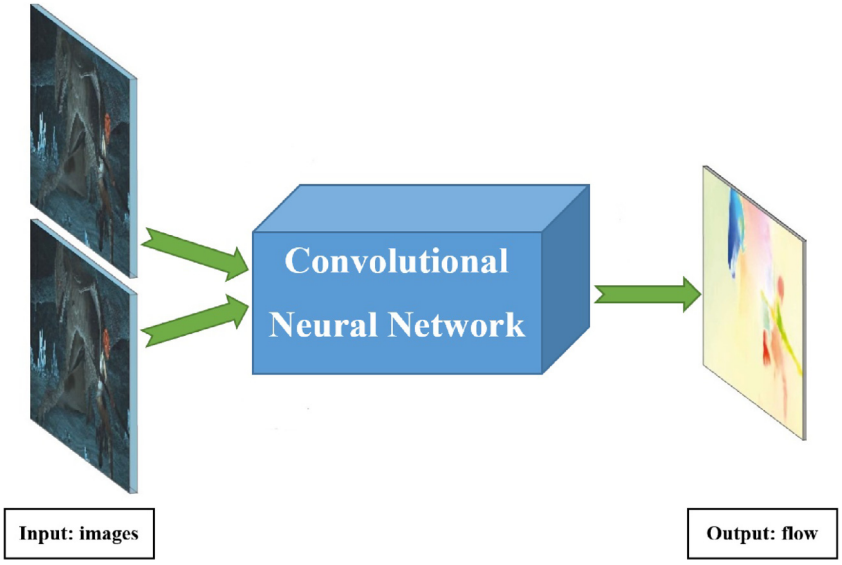

Fig. 2. End-to-end CNN learning model for optical flow estimation.

Table 1
MPI-Sintel special evaluation protocol.

Method	EPE all	EPE matched	EPE unmatched	d0–10	d10–60	d60–140	s0–10	s10–40	s40+

Table 2
KITTI special evaluation protocol.

Method	Setting	Out-Noc	Out-All	Avg-Noc	Avg-All	Density	Runtime	Environment

Table 3
Characteristic comparison of the available public datasets.

Datasets	Middlebury	KITTI	MPI-Sintel	Flying Chairs
Image pairs	72	194	1041	22872
Ground truth (GT)	08	194	1041	22872
GT density per image (%)	100	~50	100	100

This error measure is defined as the root-mean-square (RMS) difference between the warped interpolation image $I_{warp}(x, y)$ and the GT image $I(x, y)$ [5].

- Normalized Interpolation Error (NIE): The NIE between a warped interpolation image $I_{warp}(x, y)$ and a GT image $I(x, y)$ is given by [71]:

$$RMS = [\frac{1}{MN}\sum_{(x,y)}\frac{(I_{warp}(x,y) - I(x,y))^2}{\|\nabla I(x,y)\|^2 + \varepsilon}]^{1/2} \quad (7)$$

Here, ε is an arbitrary scaling constant, typically set to 1. Interpolation error measures (Eqs. (6) and (7)) require a robust interpolation algorithm. In addition, they rely on the homogeneity of the scene, as any incorrect flow vector pointing to a location with an identical brightness value is considered incorrect.

1.4.2. Color-coded visualization

To more vividly perceive the performance of optical flow, color coding approaches were introduced to visualize the optical flow and the error map. Fig. 1(b) and (c) show two types of flow field. Each provides a qualitative insight of the accuracy of the estimation. The color-coded flow field (Fig. 1(b)) is a dense visualization of the optical flow field. A color hue is associated to each direction and the saturation of the color increases with the magnitude of the flow vector (see Fig. 3). The vector-plot flow field (Fig. 1(c)) directly represents the displacement vectors and provides a good intuitive perception of physical motion.

Errors in the estimation of the optical flow can also be visualized [72]. Fig. 4 shows the correct pixels are colored in blue while the error pixels are colored in red. Such visualizations provide a readily understood means of understanding the performance of an optical flow algorithm.

1.5. Evaluation datasets

Publicly available datasets with ground truth optical flow allow researchers to compare their novel algorithms to existing work, and understand relative strengths and weaknesses of the methods. There is a steady progression in the size and realism of these benchmark datasets. We briefly discuss here the four most common optical flow datasets. The characteristics of the datasets are compared in Table 3.

1.5.1. Middlebury

The Middlebury benchmark dataset [5] provides ground truth for both real and synthetic sequences with complex conditions such as motion discontinuities, non-rigid motion, motion blur, and multiple independent motion. It does not contain very large displacements.

1.5.2. MPI-Sintel

This benchmark contains long photo-realistic sequences with extremely difficult cases [73]. It includes large motion, specular reflections, motion blur, defocus blur, and atmospheric effects. For evaluation, the basic EPE and some variants derived from EPE are used. The MPI-Sintel protocol measures the error distribution with respect to occlusion and large motion by different thresholds, see Table 1.

1.5.3. Flying Chairs

The Flying Chairs datset [63] is sufficiently large to train CNNs for optical flow computation. The dataset consists of 22,872 image pairs with corresponding optical flow fields. Images are collected from Flickr on which the segmented images of chairs from [74] are overlaid. Random affine transformations are used to chairs and background to acquire the second image and ground truth optical flow fields.

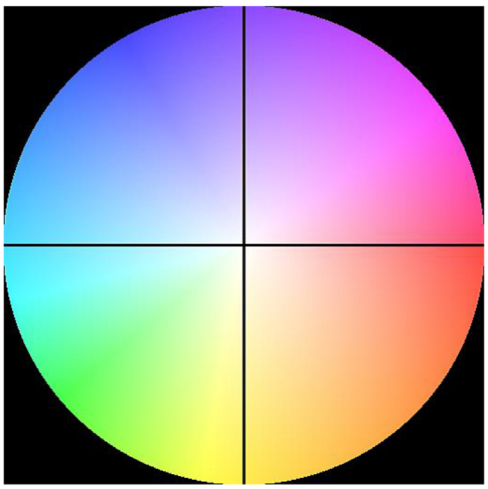

Fig. 3. HSV color space: Direction is coded by hue, length is coded by saturation.

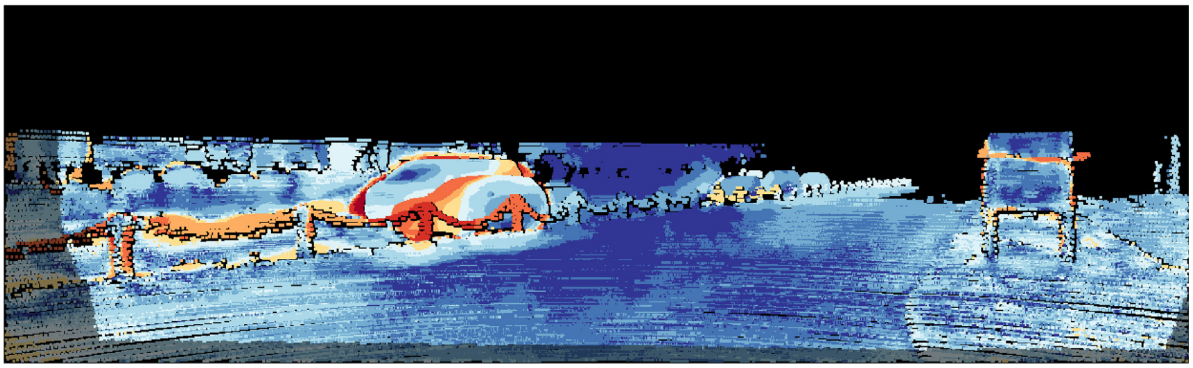

Fig. 4. A flow error map colorization in KITTI 2015 dataset.

1.5.4. KITTI

The KITTI benchmark [75] consists of real sequences taken from a driving platform in an uncontrolled environment. The scene is much more challenging compared to other benchmarks because it includes non-Lambertian surfaces, different illumination conditions, a large variety of materials and large motion. Recently, background and foreground annotations were provided for the dataset [72]. A specific evaluation protocol was introduced for the KITTI benchmark. It uses an EPE threshold of τ pixels ($\tau \in (2, \ldots, 5)$), and computes the percentage of pixels whose EPE is above the threshold. With occlusion ground truth, the metric is shown in Table 2.

2. Variational optical flow technique

In 1981, Horn and Schunck (HS) [1] proposed to compute the displacement vector of every pixel by minimizing a global energy function that is a combination of a BCA-based data term and a smoothness term. A smoothness parameter λ was to balance the two terms. Many extensions and modifications of this variational optical flow technique have been proposed [35,76,77]. According to the most recent survey [5], the variational method plays a dominant role in flow computation since almost all top-performing flow algorithms are based on the variational framework [77].

There are three primary advantages of the variational optical flow method when compared to other kinds of optical flow techniques [6]:

- It can integrate different assumptions into a single minimization framework.
- It has the filling-in effect which yields a dense flow field, whereas many other techniques require post-processing to interpolate the sparse flow field.
- Its energy function can be formulated in such a way that it is invariant under rotations in most cases.

Additionally, the variational method can generally be implemented to run in real-time using modern numerical approaches [48], or advanced computation techniques (FPGA [78] or GPU [49,79]). We will now introduce the classical variational method and its most important improvements.

2.1. The classical variational method

The fundamental principle behind the variational optical flow method is the BCA, which assumes that the brightness of a moving pixel remains constant over time. Mathematically, it can be formulated as:

$$I(x, y, t) = I(x + u, y + v, t + 1) \tag{8}$$

Where $I : (\Omega \times T \rightarrow \Re)$ is a image sequence, $I(x, y, t)$ is the current frame at time t, $I(x + u, y + v, t + 1)$ is the next frame at time $t + 1$. $(\Omega \rightarrow \Re^2)$ denotes the image domain and T is the sampled time interval of the sequence. $\mathbf{x} = (x, y)$ represents a point in the image domain Ω. $\mathbf{w} = (u, v)$ denotes the flow field, with u and v the horizontal and vertical flow component, respectively.

Eq. (8) produces a difficult optimization problem. HS applied a first-order Taylor expansion to linearize the right-hand side, which yields the approximation:

$$u\frac{\partial I}{\partial x} + v\frac{\partial I}{\partial y} + \frac{\partial I}{\partial t} = 0 \tag{9}$$

This is the well-known *optical flow constraint* (OFC). Eq. (9) can be reformed as:

$$\nabla I(\mathbf{x})\mathbf{w}(\mathbf{x}) + I_t(\mathbf{x}) = 0 \tag{10}$$

Based on the BCA, the *data term* is formulated as:

$$E_D(u,v) = \int_\Omega (I(x,y,t) - I(x+u, y+v, t+1))^2 d\Omega \tag{11}$$

Only the BCA based constraint leads to an undetermined equation system. To overcome the ambiguity, HS proposed a smoothness constraint which dictates that the flow field should vary smoothly as neighboring pixels tend to have similar motion. In other words, the flow variation between a point (x, y) and its neighbors are close to zero. Mathematically, it can be expressed as:

$$|\nabla u|^2 + |\nabla v|^2 = (\frac{\partial u}{\partial x})^2 + (\frac{\partial u}{\partial y})^2 + (\frac{\partial v}{\partial x})^2 + (\frac{\partial v}{\partial y})^2 \tag{12}$$

According to the smoothness assumption, the *smoothness term* is defined as:

$$E_S(u,v) = \int_\Omega (|\nabla u|^2 + |\nabla v|^2) d\Omega \tag{13}$$

A global energy function for optical flow is formulated by combining the *smoothness term* with the *data term*:

$$E(u,v) = \int_\Omega \underbrace{(I(x,y,t) - I(x+u, y+v, t+1))^2}_{Data\ Term} d\Omega + \lambda \int_\Omega \underbrace{(|\nabla u|^2 + |\nabla v|^2)}_{Smoothness\ Term} d\Omega \tag{14}$$

Here, λ is the parameter that controls the balance between the two terms.

2.2. Modifications and extensions

The effectiveness of the HS model is constrained to constant illumination, and Lambertian surfaces with full visibility. Realistic videos typically do not meet these requirements. For example, HS is not capable to handle occlusions causes by multiple moving objects [80,81] to deal with motion discontinuities [38]. Also, illumination changes and motion blur violate the BCA and smoothness assumption [82–84]. When displacements are larger than the object structure, matching typically fails due to the limitations of the Taylor expansion and the coarse-to-fine strategy [38,85,86]. Finally, different types of video frames might have different qualities and a fixed smoothness parameter is, therefore, an unsuitable factor to balance the data term and the smoothness term [87].

Subsequent improvements have been dedicated to remedy the drawbacks of the classical variational method. We discuss the three main categories of improvements subsequently.

2.2.1. Data term improvement

HS utilized a quadratic penalization on the BCA to model the data fidelity. In practice, the BCA can be easily violated due to noise, illumination changes and occlusions. The quadratic penalization function lacks the capability to handle these situations. Various kinds of methods have been proposed to improve the performance of the data term.

1. Gaussian filtering has been introduced as a preprocessing step for variational methods to improve the performance of the data term with respect to noise [35,36,82].
2. To address illumination changes, three schemes have been proposed:
   - **Robust constancy constraint**. Brox et al. [35] imposed a gradient constancy term into the data term. Papenberg et al. [85] found that other high-order constancy constraints such as the constancy of the Hessian and the constancy of the Laplacian, are also useful. Zimmer et al. [88] used the gradient constancy together with the brightness constancy in the HSV color space to handle illumination changes. Xu et al. [77] pointed out that using either gradient constancy or brightness constancy was more suitable than using both. They presented a binary weight map to switch between the two terms. Generally, higher-order constancy constraints are invariant under additive illumination changes but not invariant to multiplicative illumination changes. Thus, they may not be effective with respect to realistic illumination changes that usually contain a multiplicative part [89]. Alternatively, Mohamed et al. [90] proposed an illumination-robust constancy based on local texture features to form a texture constancy constraint for the data term.
   - **Structure-texture decomposition**. Inspired by the idea of [91], to gain robustness against illumination changes, Wedel et al. [92] applied the Rudin–Osher–Fatemi (ROF) technique [93] to decompose the input image into “structure" and “texture" components, and to reconstruct the new image as a linear combination of both with emphasis on the “texture" component. Additive illumination changes are sensitive to structures while the texture part is less affected. A main defect of this method is the potential loss of valid information. Especially when no brightness variations occur, this method removes useful information of the structure component.
   - **Color space**. This kind of method aims at handling illumination changes using photometric invariant color spaces. The robustness during brightness variations is achieved by introducing a color channel transformation. The input color images are transformed to a color space where the channels are invariant to illumination changes. The photometric invariants of the HSI color space [89,94], the normalized RGB channels [94], the spherical space [89], and the HSV space [82,83] have been investigated.
3. Robust penalty functions. The quadratic penalty function used by HS severely penalizes larger displacement errors. To reduce the impact of such local errors that are considered as outliers, robust penalty functions were proposed [38,95]. Common used alternatives are the Charbonnier [36,96], generalized Charbonnier [76], Tukey [97], Lorentzian [38], Leclerc [98], and modified Hampel penalty function [99]. In a recent survey paper, Sun et al. [100] suggest to use the Charbonnier or slightly non-convex penalty function. On the other hand, the robust penalty function causes nonlinearity to the data term. The graduated non-convexity (GNC) strategy [101] aids in improving the performance in this case.

2.2.2. Smoothness term improvement

The smoothness term aims at smoothing the motion field in regions of coherent motion while preserving discontinuities at motion boundaries. To improve the performance of the smoothness term in discarding outliers, preserving discontinuities and propagating information in low texture areas, numerous improvements have been proposed. A taxonomy of smoothness terms appears in [102].

- **Image- and flow-based terms**. The homogeneous smoothness term proposed in the original HS model does not consider flow discontinuities and over-smoothes motion boundaries. Isotropic and anisotropic image-driven smoothness terms to suppress smoothing across image boundaries are introduced in Alvarez et al. [86] and [103], respectively. Schnorr [104] designed an isotropic flow-driven smoothness term that also avoids over-segmentation of strongly textured structures. Weickert et al. [102] constructed an anisotropic flow-driven smoothness term that achieves smoother

effects along flow discontinuities while producing fewer fluctuations. Zimmer et al. [88] exploited a joint image- and flow-driven smoothness term to avoid artifacts of over-smoothing and over-segmentation. Recently, the advanced motion and structure-adaptive regularization term is widely used for preserving edges [77,88]. This term favors motion discontinuities to coincide with discontinuities of the image structures.

- **Non-local regularization**. The gradient of the flow can only supply a local constraint on the interaction between pixels. Large-range interactions can capture the form of the displacement field more accurately. According to this assumption, non-local smoothness terms have been recently studied in [76,105,106] by describing the structure of the flow in an extended neighborhood.
- **Spatio-temporal regularization**. Assuming the flow field varies smoothly and gradually over time, the spatial-smoothness term can be extended to the temporal domain. Similar to the spatial case, temporal smoothness can be achieved locally either in terms of the temporal flow gradient [39,82,95], or by taking into account temporal coherence by modeling along the trajectory of the moving object [107,108].

2.2.3. Smoothness parameter improvement

Nagel and Enkelmann [109] suggested to make λ a function of the iteration number, but this proved ineffective. An approach with the blurring operator for choosing λ in ill-conditioned inverse problems has been described in [110]. Ng and Solo [111] replaced the blurring operator by the Euclidean distance. These brute-force methods that search for the smoothness parameter are computationally expensive. Zimmer et al. [82] presented an optimal prediction principle (OPP) to automatically determine the optimal λ, but it is limited to constant speed and linear motion. Different from these methods which aim to find a globally optimal λ, are local fusion methods that estimate the appropriate λ for each point independently. Raket [112] fuse flows of different smoothness parameters. The local evaluation in terms of the best data fit on the gradient image is easily affected by illumination changes, and is sensitive to noise and discontinuities. Tu et al. [113] address this issue evaluating the local interpolation error in terms of weighted L1 block match on the corresponding set of intensity images.

3. Challenges

Despite considerable progress over the past decades, there are still unsolved challenges for optical flow methods. These include dealing with occlusions, preserving motion boundaries, dealing with large displacements, removing outliers, handling textureless areas and rotation. We discuss these challenges below.

3.1. Occlusions

Occlusion is a common phenomenon when dealing with realistic scenes. One view is that occlusion is produced due to layers that ordered in depth [114]. A point is then occluded if it switches from one motion layer to another between consecutive frames [115]. Often, occlusions are an explanation of optical flow mismatching [80,115]. Violations of the BCA of optical flow accumulate in occluded areas [116]. A brightness change between corresponding points in subsequent images denotes an occlusion of the point under the assumptions of Lambertian reflection and constant illumination [117,81].

Occlusion can be classified into three categories. First, self occlusion occurs when one part of an object occludes another. Second, inter-object occlusion occurs when two moving objects occlude each other. Finally, background occlusion occurs when a structure in the background occludes moving objects.

To increase the robustness of the data term to occlusions, Black and Anandan [38,95] replaced the quadratic function with a robust penalty function. Also, some more effective regularizers have been presented [82,85,102,104,118]. These methods aim to reduce the errors caused by occlusions, but still fail when pixels fully disappear. Recently, researchers suggested to explicitly deal with occlusions in a two-steps manner. First, to detect occlusions and then to process the occluded areas [76].

3.1.1. Occlusion detection

Occlusion detection is crucial in the estimation of optical flow as it is easier to compute optical flow if the occluded regions were known. Many methods have been proposed to detect occlusions. Some approaches analyze the mismatch between the forward and the backward flow to detect occlusions [80,119,120]. Others use the residual from optical flow estimation to determine whether a region is occluded [81,121,117]. Some approaches exploit features of the optical flow field, such as edges [122,123], motion cues [124,125] and depth [125]. Confidence measures have also been proposed to detect occlusions [126,55]. A promising avenue for improvement is in the combination of optical flow features such as motion constancy [127] with image features (e.g. brightness, color, texture, depth) [81,124,127–129].

3.1.2. Detailed introduction and analysis

Thompson [130] applied a boundary detection scheme to classify boundaries into an occluding group and an occluded group. This distinction is then used to alleviate the impact of occlusions by a flow projection strategy. Alvarez et al. [80] checked the consistency between the forward and backward flow with a threshold to detect occlusions, and then ignored the occlusions by resetting the detected pixels to zero in the post-processing phase. Xiao et al. [117] modified [80] by reformulating the energy function by explicitly introducing an occlusion term to balance the energy loss due to occlusion. A multi-cue driven bilateral filter was exploited to substitute the original anisotropic filter. This novel filter can adaptively control the diffusion process according to the detected occlusion information and can thus avoid diffusing flow across occlusion boundaries. Sand and Teller [121] improved [117] by utilizing an occlusion-aware bilateral filter, which combines the flow divergence and pixel projection difference to detect occlusions. Occlusions were tackled by deleting and creating particles in disoccluded areas. Ince et al. [120] proposed a variational formulation to jointly compute optical flow, implicitly detect occlusions and extrapolate optical flow in occluded areas. This approach permits interaction between optical flow estimation and occlusion detection. Ayvaci et al. [81] treated residuals of occlusion and non-occlusion separately. Rua [115] exploited a local spatio-temporal image-reconstruction model that only needs the information of a plausible motion captured from the image pair for occlusion detection. In this manner, occlusions can be more directly identified without the requirement of a precomputed dense motion field. Zhang et al. [131] proposed to detect occlusions by comparing the brightness changes of the triangulation and embedded pixels between successive frames.

3.2. Over-smoothing

Imposing smoothness is one approach to cope with the aperture problem but comes at the cost of smoothing motion boundaries. Techniques have been presented to allow piecewise smoothing while preserving discontinuities by improving the regularization term to better fit the flow field. In addition to the measures discussed in Section 2.2.2, we summarize several other effective approaches.

3.2.1. Statistical learning-based approaches

Learning the statistics of the flow field and utilizing it as a prior to regularize the flow field can improve the overall flow accuracy because prior information is useful to reduce smoothness in textured regions and discontinuities. Roth and Black [45] learned spatial smoothness by using Field-of-Experts. Sun et al. [118] learned statistical models of both data constancy error and image structure-adaptive flow derivatives. Shen and

Wu [132] introduced a sparsity prior to calculate optical flow. This novel measure is effective in preserving motion discontinuities because these can be accurately estimated by finding the sparsest representation of the flow field. Jia et al. [133] extended this approach into a learned sparse model (LSM), which obtains combines first-order spatial regularity and the learned, higher-order, sparse model. Moreover, it does not need separate regularization of smooth motions and motion discontinuities.

3.2.2. Segmentation-based approaches

Coupling optical flow estimation with segmentation can offer better support for the preservation of flow discontinuities. Decomposing the image sequence into coherently moving objects or different layers would greatly simplify the optical flow solution. This is because it is able to work in regions that do not contain discontinuities, and also motion boundaries are avoided. Memin et al. [134] presented a simple mechanism that enables interaction between a dense discontinuity-preserving estimation process and a segmentation process. The joint estimation-segmentation model introduces a mixture between local smoothness and region-wise parameterization, which is beneficial for flow field discontinuity preservation. Amiaz and Kiryati [135] proposed an algorithm that embeds the energy function of [85] within an active contour segmentation framework. It obtains the accurate performance as [85] while producing sharper flow boundaries. Xu et al. [136] introduced a modified segmentation-embedded optical flow framework that accommodates the parametric and segmented motion estimation in a variational model to handle discontinuities. Lara et al. [137] designed a novel model of localized layers to enhance the performance of optical flow computation with respect to semantic image segmentation. By using layered optical flow in constrained regions around objects of interest, the localized layer model overcomes the drawback of the typically global layered model which is unable to represent complex motion boundary relationships.

3.3. Large displacements

For the variational algorithm, the data term is usually linearized by Taylor expansion. This approximation performs badly when dealing with large non-linear displacements because the linearization is only valid for velocities of limited magnitude [92,138]. Additionally, there is the increased risk of getting trapped in a local minimum when the displacement increases. We introduce techniques to handle this problem in three aspects.

One common way to deal with large displacements is to embed the estimation in a coarse-to-fine or warping framework [38,52,98,85,139]. In such a framework, downsampled pyramid images are produced. A flow increment is estimated between the first image and the warped second image at the current scale. Then the accumulated flow field is propagated to the next finer level until the finest level is reached [35, 76,140]. Anandan [141] introduced the unidirectional coarse-to-fine scheme to the variational model to calculate dense flow fields. Starting with Black and Anandan [38], the unidirectional coarse-to-fine strategy combined with other techniques became more widely used in variational algorithms [49,76,92]. Bruhn et al. [48,142] and Zimmer et al. [82] proposed an advanced coarse-to-fine warping technique: bidirectional multigrid scheme. This strategy is useful but also introduces some issues:

- The interpolation to resize image and flow introduces errors. Larger artifacts are produced on the next finer scale when upsampling flows at occlusion boundaries [34].
- Selecting a suitable downsampling factor is non-trivial but affects the overall-performance of the algorithm to a large extent [86].
- Oversmoothing of fine structures and failure to capture small-scale and fast-moving objects [143] is a significant risk. Downsampling reduces the size of the image and the displacements. If the object size is smaller than its displacement, it is likely to be smoothed out at coarse level and will not be recovered accurately (see Fig. 5).

To remedy these drawbacks, Alvarez et al. [103] did not conduct linearization to allow for larger displacement. A linear scale-space focusing scheme was exploited for avoiding convergence to incorrect local minima. This method only partially handles the large displacement problem as it still depends on sub-sample warping. Steinbruecker et al. [138] presented a strategy to avoid both linearization and warping. Their algorithm requires an exhaustive search for pixel-level candidate matching, which is computationally expensive and makes the estimation intractable.

Recently, there has been a great interest in using feature matching to address the issue of large displacements. Two main directions have been exploited in literature.

3.3.1. Integrating feature matching into the energy function

The SIFT-flow method [144] employed SIFT [145] descriptors to compute a scene flow field. Since this method fully depends on SIFT descriptors, it has difficulty estimating small-motion regions [146]. Inspired by [147], Brox et al. [3] added a feature matching constraint into the classical energy function. In this way, correspondences from descriptor matching are obtained, which are helpful to support the coarse-to-fine strategy in avoiding local minima. This method has some defects, for example, the local descriptors are reliable only at salient locations, and false matches are typically generated. To handle these issues, subsequent modifications of [3] have been proposed. Stoll et al. [148] presented an adaptive integration strategy to reduce false point matches. Zin [149] applied segment matching [150] to modify the matching component. The matching term is modified to deal with weakly localized line features. Weinzaepfel [146] proposed a DeepFlow algorithm which involved dense and deformable matching, to gain performance for fast motion. Revaud et al. [151] improved the performance of [146] by replacing the HOG descriptor in the matching term of [3]. Their DeepMatching algorithm is able to handle non-rigid deformations explicitly. In addition, it is robust to repetitive or weak textures by incorporating a multiscale scoring of the matches. Since DeepMatching is based on gradient histograms, it is insensitive to appearance changes that arise from illumination and color variations. Kroeger et al. [152] applied inverse search for fast patch correspondences. This produces a dense displacement field in terms of patch aggregation along multiple scales.

3.3.2. Fusing matching information with variational flow

Lempitsky et al. [46] is among the first to apply fusion to improve optical flow estimations. They proposed to fuse flow candidates obtained from different flow methods or a single flow method with different parameter settings. Derived from this idea, sparse matches and approximate nearest neighbor fields (ANNF) have become popular.

Xu et al. [77] applied sparse SIFT matching and PatchMatch [153] to compute flow candidates. When fusing them with the initial sub-pixel dense optical flow, much more accurate results can be obtained. Revaud et al. [154] used a sparse set of matches to construct a dense correspondence field by performing a sparse-to-dense interpolation. This method, EpicFlow, is not only fast but it is also robust to motion boundaries and large displacements. However, it is vulnerable to input matching noise. To deal with this issue, Hu et al. [155] proposed a robust interpolation method. By assuming the flow in each superpixel satisfies a piecewise affine model, the optical flow can be robustly estimated. Li [156] exploited a pyramid gradient matching method to produce robust dense matches for optical flow computation. Due to the efficient and scalable nature of the gradient image, the method is fast and accurate.

ANNF has been used by Chen et al. [139] to compute an initial motion field which contains different types of correspondence information, and then fuse these with variational flow candidates for refinement. The use of ANNF comes with two issues. First, there are usually many outliers that are hard to identify because there is no regularization.

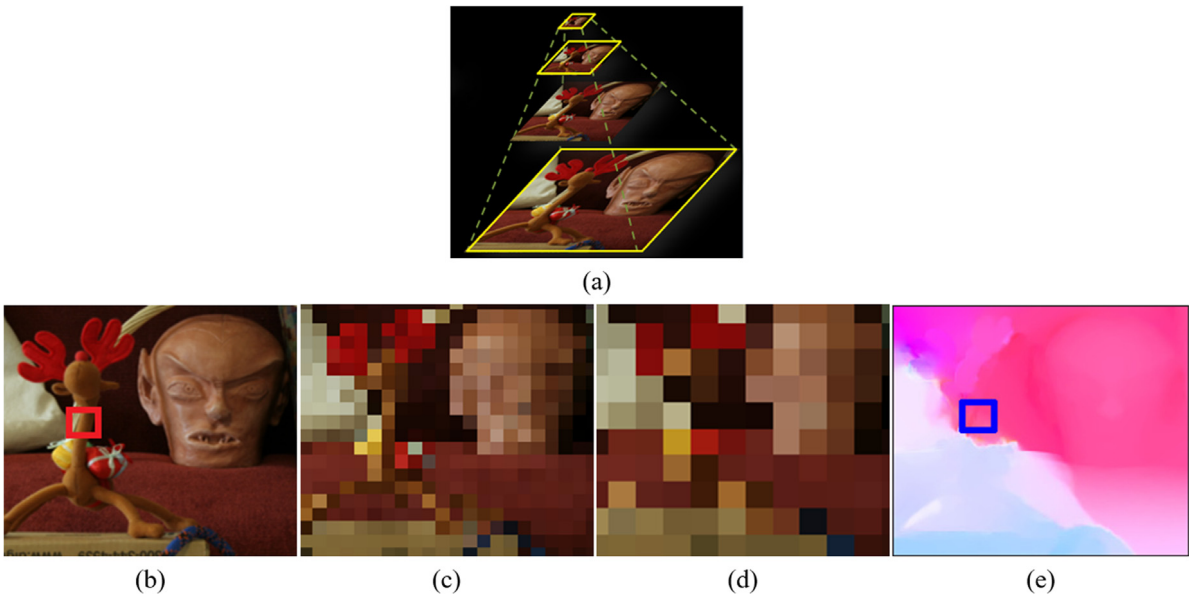

Fig. 5. (a) The coarse-to-fine strategy. (b) Original image. (c) Third level image. (d) Fourth level image. (e) Estimated flow. Note that the neck is not recovered. Image taken from [77].

Second, ANNF is computational inefficient. There have been several modifications to remedy these issues.

Lu et al. [157] exploited a generic PatchMatch filter method to handle multi-labeling problems efficiently using fast superpixel-based search strategies. Not only the computational cost is significantly reduced but also accurate optical flow is obtained as the generated ANNF is edge-aware. Bailer et al. [158] presented a novel purely data based hierarchical correspondence field search strategy which is able to not only find most inliers but also to reject outliers effectively. Bao et al. [79] exploited a fast randomized edge preserving ANNF. By propagating self-similarity patterns similar to [159], the computational time is significantly decreased. Hu et al. [160] proposed CPM (Coarse-to-fine PatchMatch) to speed up the computation by combining an efficient random search strategy with a coarse-to-fine scheme. As a propagation step with constrained random search radius between successive levels on the pyramid architecture was used, outliers are filtered and the correspondences are smoothed.

How to optimally fuse the matching information with the variational flow is the topic of ongoing research. The common fusion method, quadratic pseudo-boolean optimization (QPBO) [161,162], has some drawbacks. The submodularity constraint imposed on the pairwise terms is hard to fully meet, and the constraint is not suitable for variational optical flow models. Tu et al. [113] proposed a weighted local intensity fusion (WLIF) approach to evaluate the local interpolation error with respect to L1 block match on the corresponding set of images. An adaptive weight based on selective gradient magnitude is exploited to handle outliers and occlusions. However, WLIF is not robust to occlusions.

3.4. Outliers

Outliers, in both the images (e.g. image noise) and the intermediate flow fields (e.g. flow errors [38]) directly affect the accuracy of the final optical flow. One flow outlier is able to influence its neighboring flow vectors due to the smoothness operation and the derivative computation. Even worse, these errors will be accumulated and diffused during numerical iteration. There are three major types of techniques aim to deal with this problem. One technique that uses penalty functions to increase the robustness to outliers, has be introduced in Section 2.2.1. In the following, we focus on the other two techniques: denoising and joint flow computation and outlier removing.

3.4.1. Denoising

Variational algorithms rely on spatiotemporal derivatives, but this process enhances noise. Fermuller et al. [163] pointed out that noise causes a bias to underestimate not only the lengths but also the directions of flow vectors. Therefore, denoising plays a crucial role in optical flow estimation [164]. We divide denoising approaches into three categories.

- **Image denoising**. Filters are applied to the input images to remove noise and other high frequencies, while favorable structures of the images are enhanced. Generally, denoising can be achieved by averaging pixels locally or non-locally.
  Gaussian filtering is a basic local linear filtering scheme. The image is convolved with spatial or spatiotemporal Gaussian kernels with a predefined standard deviation [1]. Weber et al. [165] modified the method to convolve the image with a set of linear, separable spatiotemporal kernels. The Laplacian filter was used by Lempitsky et al. [166]. Song et al. [167] proposed a two-step denoising approach. In the first step, the Gaussian filter is employed to reduce high-frequency noise and in the second step, a generalized regularization method is applied for deblurring. Due to the high efficiency and edge-aware smoothing of guided image filtering (GIF) [168], Tu et al. [169] exploited a novel adaptive guided image filter to correct errors caused by outliers in the warped interpolation image during numerical computation to boost the accuracy of optical flow estimation.
  Buades et al. [170] proposed a non-local filtering approach to remove noise by averaging pixels in a non-local area. Liu and Freeman [171] designed an adaptive denoising framework that incorporates robust optical flow into a non-local form to remove incidental and structured noise introduced by digital cameras.
- **Flow denoising**. Denoising the flow field is an effective way to improve the performance of optical flow computation in three different ways. First, it removes noise and outliers in the flow field. Second, it preserves flow edges without over-smoothing. And finally, it propagates valid flow information into low texture areas. We discuss several techniques.
  - *Median filtering (MF)*. With MF, outliers can be well discarded [76,92,172,173]. Sun et al. [76] formulated the MF heuristic as a non-local term [170] in the objective function, and modified the performance by integrating information about image structure and flow boundaries into a

weighted non-local term (a weighted median filter (WMF)) to prevent over-smoothing. The color similarity measure, the most powerful cue of the WMF, can be easily perturbed by noisy pixels. To increase the robustness of the WMF to noise, Tu et al. [173] presented a PatchWMF method, which introduces the idea of non-local patch denoising to compute the color similarity in terms of patch difference. In addition, they designed an improved color patch similarity measure.

- *Bilateral filtering (BF)*. BF is a non-linear, edge-preserving and noise-reducing smoothing filter. The output of each pixel is calculated as a weighted average of intensity values in a local neighborhood [174]. Xiao et al. [117] replaced the traditional anisotropic diffusion process with a multi-cue driven BF and separated the variational method into two update stages. Similar to MF, [117] is able to adaptively control the diffusion process according to the detected occlusion information, image intensity and motion dissimilarity. However, this method is computationally expensive. Sand and Teller [121] improved its efficiency by applying the BF only near flow boundaries. Tu et al. [140] proposed a combined post-filtering method that first classifies the flow field into edges, occlusions, and flat regions. Each region is smoothed using a weighted median filter (WMF), bilateral filter (BF) and fast median filter (MF), respectively. A hybrid gradient bilateral filter and Gaussian filter (HGBGF) are exploited to further solve the low efficiency issue of BF.
- *Kalman filtering (KF)*. KF allows for the estimation to be incrementally updated with each measurement. The velocity measurements of each frame can be integrated with the previous ones and uncertain measurements can be attached to the velocity of each point. This information is useful to reduce outliers. Additionally, more accurate estimation can be achieved by repeating the measurements [175]. Recently, Rabe [176] combined KF with the variational approach to establish a temporal smoothness of the dense 3D motion field.

- **Joint flow computation and outliers removing**. In this technique, outliers are explicitly considered in the energy function and handled during optimization. By parameterizing the appearance of each frame as a function of both the pixel motion and the motion-induced blur, Portz et al. [177] improved the data term to compute optical flow in the presence of spatially varying motion blur. Tu et al. [84] modified the efficiency of [177] by designing a down-sample interpolation technique during the blur detection, and applied an edge-preserving regularization approach. Recently, Tu et al. [178] proposed a new hybrid objective energy function that jointly computes optical flow and restores images with preserved edges by alternately minimizing an optical flow module and a denoising restoration module. These methods face one common issue that deblurring or denoising the images during optimization is not always beneficial for boosting the performance as some useful details would be removed.

3.5. Textureless

The homogeneous regularization term of the HS model leads to flow vectors with constant smoothing over the flow field. However, if the images contain regions with weak textures, the variational methods fails to produce correct optical flow due to the poor image derivatives information [179].

Two kinds of techniques are introduced to handle the issue of textureless. First, the regularization term can be improved. In [95,102], the regularization term was modified by integrating temporal information. Trobin et al. [180] proposed a regularization term based on the decorrelated second-order derivatives to avoid the total variation (TV) induced staircase effects, especially on weakly textured regions. Xu et al. [136] regularized affine parameters and the segmentation information into a variational model. Muller et al. [181] exploited a modified TV regularization term with stereo and feature information. Palomares et al. [182] introduced a Non-Local Total Variation (NLTV) regularizer to incorporate low-level image segmentation in a unified variational architecture. In this manner, motion from textured areas is propagated to the untextured regions.

The second approach to improve optical flow estimation in weakly textured regions is by integrating feature term. Revaud et al. [151] introduced DeepMatching into the matching term of the variational energy function. DeepMatching is robust to weak textures because it uses a multiscale scoring of the matches. By considering patches at multiple scales, the lack of distinctiveness that affects small patches can be overcome.

3.6. Rotation

Image rotation is common but is a challenging problem in optical flow estimation. Three kinds of methods deal with this problem:

- **Modified Derivatives**. Numerical methods to calculate image derivatives lack rotational invariancy. Freeman and Adelson [62] proposed a steerable filters technique to compute the directional derivatives in a straightforward way. Niu [183] presented a modified strategy to compute the directional derivatives based on their original definition. Aubert and Kornprobst [184] computed the gradient with neighbors and applied a weighted factor to control the significance of the directional derivatives.
- **Modified coordinate systems**. Smoothness constraints regularize each pixel's flow by its neighboring flow vectors. However, the horizontal–vertical coordinate system often fails to represent the intrinsic image structures. For example, a pixel and its horizontal and vertical neighbors have a different velocity during rotation. Various coordinate systems were constructed to handle this problem. Ho and Goecke [185] transferred the horizontal–vertical coordinate system into the log-polar coordinate system. Niu et al. [186] designed an adaptive local structure-oriented coordinate system.
- **Modified constraints**. Since the traditional data and smoothing terms do not consider the rotation, additional constraints can be integrated into the function. Liu et al. [144] used SIFT, which is rotation and scale invariant, into the data term to handle rotation. Other scale invariant descriptors [187,188] are also exploited.

4. CNN-based technique

CNN-based Technique learns to extract deep features from input images. For a novel test image pair, optical flow is then calculated by an optimization of the learned features. Inspired by the recent success of deep convolutional neural networks (CNNs) in many computer vision and image processing tasks, Dosovitskiy et al. [63] were among the first to present a CNN architecture called FlowNet to learn optical flow directly from synthetic annotated data. Its efficiency is several orders of magnitude higher than the variational method. This landmark achievement encouraged subsequent supervised, unsupervised, and semi-supervised CNN-based methods for optical flow estimation. Currently, CNN-based algorithms provide a promising alternative to the variational method. Because its very different nature, we first outline the major differences between the two classes of technique. We then discuss supervised, unsupervised and semi-supervised methods in detail.

4.1. Differences between CNN-based and variational methods

We compare CNN-based and variational methods by discussing the relative advantages and disadvantages of CNN-based methods.

One of the main advantages of CNN-based approaches is that they are more flexible in the image features they use for optical flow estimation. Multi-layer and hierarchical architectures enable CNNs to extract more abstract, deeper, and multiscale features. Also, CNNs can model complex, non-linear transformations between the input images and the estimated flow field. Overall, the stochastic minimization of the loss across an entire training dataset avoids some of the pitfalls of optimizing a complex energy-function on individual inputs in variational methods [189]. As a result, fewer assumptions on the input data need to be introduced. Importantly, CNN-based optical flow algorithms are must faster than variational methods.

CNN-based methods also have several drawbacks. Given that the parameters of CNNs are learned from training data, the performance of the CNN-based technique heavily depends on the quality and size of this labeled training dataset. It is difficult to obtain dense ground truth labelings for real scenes. Therefore, datasets with synthetic, rendered, sequences have been introduced. While these contain realistic motions, they do not reflect the complexity of realistic photometric effects such as image noise, illumination changes, shading and motion blur.

CNNs contain many parameters, typically in the order of millions. This means that there is a significant risk of overfitting, especially given the difficulty of obtaining suitable training data. Also, this causes CNNs to have a large memory footprint. The large number of parameters also hinders intuitive understanding of the inner workings of the network [190]. Finally, several hyperparameters significantly affect the efficacy and efficiently of the learning process. For example, it is hard to set a good loss function.

4.2. Supervised methods

Supervised methods are the predominant way of training CNNs capable of optical flow estimation. Their performance, both in terms of accuracy and efficiency, is typically good. However, supervised methods require a large amount of ground-truth optical flow to train the parameters of the CNNs. Also, there is no guarantee that the trained models generalize to different scenarios. Here, we distinguish between end-to-end methods that perform both feature extraction and matching, and methods that only perform one of these tasks.

4.2.1. End-to-end

Dosovitskiy et al. [63] exploited an end-to-end CNN architecture, FlowNet, to compute optical flow. Two CNN networks, FlowNetS and FlowNetC, are constructed based on the U-Net denoising autoencoder. The CNNs are trained in a supervised way on the FlyingChairs dataset they published. The optical flow estimation can be performed efficiently, but the quality of the results are low and the network is constrained for a specific image resolution. Moreover, the EPE loss function that is used in the training procedure ignores the relation between image pairs and optical flow. Based on the architecture of [63], Mayer et al. [191] constructed a more sophisticated synthetic dataset with 3D scene information to train CNNs for optical flow estimation.

Many adaptations and extensions of FlowNet [63] have introduced to reduce the size of the trained CNN model, improve the running efficiency, and enhance the accuracy of the optical flow estimation. By taking into account the constraint relationship between image pairs and optical flow during training, Xiang et al. [192] exploited an additional supervised term in their loss function. Ranjan and Black [193] designed a Spatial Pyramid Network (SPyNet) following the coarse-to-fine idea of the variational method to deal with large displacements. At each pyramid level, a convolutional network is learned to compute optical flow, which is then up-sampled to the next level to guide the warping of the second image towards the first. Compared to FlowNet, there are significantly less model parameters and the results are more accurate. However, its accuracy is lower than FlowNetS [63]. Extending FlowNet by stacking multiple encoder–decoder networks resulted in FlowNet 2.0 [67]. This network reduces the estimation error by more than 50%. The success is partly due to the introduction of a subnetwork focused on small, subpixel motion. FlowNet 2.0 has three limitations. First, its subnetworks need to be trained sequentially to reduce over-fitting. Second, the model size is large (over 160 M parameters) which is not suitable for mobile and embedded devices. Third, FlowNet 2.0 is slower than FlowNet.

Inspired by FlowNet [63] and the coarse-to-fine pyramid strategy, several more effective methods are proposed [194,195]. Sun et al. [194] proposed PWC-Net, which is based on simple and well-established principles such as pyramidal processing, warping, and cost volume. In contrast to FlowNet 2.0, the model size of PWC-Net is about 17 times smaller, leading to an application real-time. Hui et al. [195] improved the current CNN frameworks in other two aspects. First, they exploit a more effective cascaded flow inference strategy. Second, a flow regularization layer is proposed to reduce outliers and compress vague flow boundaries. Zhao et al. [196] presented a Multi-Scale Correspondence Structure Learning (MSCSL) approach to jointly model the multi-scale correspondence structure in a flexible architecture using a Spatial Conv-GRU neural network based on multi-level deep features.

To handle the black-box nature of the CNN-based methods, and to understand the trained network is/is not suit to work on which cases, Ilg et al. [190] studied methods to estimate the uncertainty of deep regression networks for optical flow. On contribution is that they try to answer an open question which approaches can be used for uncertainty estimation. The other contribution is that they exploit a multi-hypotheses network produces multiple hypotheses in one single-network without the need of sampling or ensembles.

4.2.2. Not end-to-end

We distinguish between CNNs used to learn deep features from the input images and CNNs employed to predict optical flow with the acquirements of pre-processing and/or post-processing. In the first category, Gadot and Wolf [197] used a Siamese CNN to compute the descriptors per pixel of each image independently. The PatchMatch method [153] is used on top of the abstracted descriptors to produce an ANNF. In contrast to the previous approaches that employed Neural Network layers to calculate the matching score, they applied the L2 metric to simplify the ANNF computation. The sparse optical flow obtained from the matches is used as input for EpicFlow [154] to obtain dense optical flow. Schuster et al. [198] improved [153] by introducing a Hinge loss to replace the DrLIM loss. Guney and Geiger [65] applied the Siamese CNN to learn per-pixel context-aware features for solving optical flow though discrete optimization. Bai et al. [199] applied two branches of a Siamese network to process subsequent images to learn features. Bailer et al. [200] proposed a multi-scale feature creation method tailored to CNNs to estimate optical flow. One significant contribution of this method is that they exploited a novel thresholded loss for Siamese networks that allows to speed up training by about two times.

Zweigand and Wolf [66] use CNNs to predict the dense optical flow. They presented a data-driven sparse-to-dense interpolation method based on a CNN without pooling to compute the optical flow. A set of sparse and noisy matches, a binary mask and an edge map are used as the network input. Hu et al. [201] proposed a RecSPy network consisting of a Siamese network to learn deep features and a Recurrent CNN that formulates the spatial pyramid as a recurrent process and is used to update and refine the optical flow at each scale of the pyramid. The refinement is based on an energy function that encodes structure and constancy constraints to improve the optical flow.

4.3. Unsupervised methods

The dependency of supervised method on large labeled datasets has recently motivated the study of unsupervised method, which is able to train CNN models on unlabeled image pairs or videos to predict optical flow. In contrast to supervised methods, the performance of the unsupervised method still falls behind but is improving.

Based on the classical optical flow constraint, Ahmadi and Patras [202] designed a loss function without regularization that is differentiable with respect to the unknown flow field. It allows the back-propagation of the error to previous layers. Yu et al. [203] replaced the supervised loss function of FlowNet by a proxy loss function, which combines a data term that imposes brightness constancy over time with a smoothness term that models the expected variation of optical flow across images, to learn optical flow end-to-end in an unsupervised way. The network is able to optimize the brightness constancy and motion smoothness assumptions explicitly by making use of differentiable warping. Since the proxy loss is too simplistic, its accuracy is much lower than FlowNet. To address this issue, Zhu et al. [204] proposed a guided optical flow learning method by combining the unsupervised proxy loss with proxy ground truth flow produced via a classical optical flow method. They applied the classical methods to produce proxy ground truth optical flow to guide the CNN training. The learned models are subsequently fine-tuned in an unsupervised manner by minimizing an image reconstruction loss. Ren et al. [64] exploited a network similar to the Spatial Transformer Network [205], utilizing the photometric consistency for end-to-end learning without supervision. The photometric error between the warped feature map from the reference image and the target image is treated as the loss, measured by the objective function used in the variational method [85].

Fan et al. [206] designed TVNet, which is formulated by imitating and unfolding the iterations in the classical variational method, TV-L1 [49], to customized neural layers. TVNet is well initialized as a particular TV-L1 approach and can obtain desirable results without additional training on ground-truth optical flows. It incorporates the strengths of both the classical variational method and CNNs. Wang et al. [207] designed a deep neural architecture to boost the performance of unsupervised flow estimation with the emphasis on addressing the issues of occlusion and large motion. First, they explicitly model the occlusion map that is caused by motion and combine it with the loss function. The occlusion problem of the common loss function, which prefers to compensate the occluded regions by moving other pixels, is handled. Second, they exploited three strategies to deal with large displacements: a new warping method to facilitate the learning of large displacements, introducing additional warped inputs during the decoder stage, and applying histogram equalization and channel representation to the flow computation. Meister et al. [189] extended the FlowNetS-based UnsupFlownet [203]. They introduced an unsupervised loss that relies on occlusion-aware bidirectional optical flow estimation. They also applied the comprehensive unsupervised loss to train FlowNetC to learn bidirectional flow. Finally, iterative refinement is conducted by stacking multiple networks of FlowNet. These methods [207,189] rely on the accuracy of the estimated optical flow, and require to set a heuristics to infer occlusions. By extending the two-frame architecture of [194] to multiple frames, Caron et al. [208] proposed a novel unsupervised formulation to jointly estimate optical flow and occlusions over multiple frames. The temporal constraints are beneficial to predict more accurate optical flow in occluded regions.

4.4. Semi-supervised methods

To benefit from both synthetic data with labeled ground truth flow and realistic data without labeled ground truth flow, semi-supervised methods have been explored [209]. One strategy is to minimize the End Point Error (EPE) loss for data with ground truth flow and the loss functions that measure the brightness constancy and spatial smoothness for unlabeled input [202]. Such a method is usually sensitive to the selection of parameters which may decrease the performance of optical flow.

Lai et al. [209] presented a generative adversarial Network (GAN) to learn optical flow with both the labeled and the unlabeled data in a semi-supervised learning framework. The adversarial loss, which plays as a regularizer for both types of data, is able to discover structural patterns of flow warp errors without making explicit assumptions on the brightness constancy and motion smoothness. Yang et al. [210] exploited a Conditional Prior Network (CPN) to predict optical flow. It consists of a prior that imposes a bias on the possible solutions and that is learned in a supervised way and a simple network that is trained to map pairs of images to optical flow in a completely unsupervised manner. The regularizer in the form of the conditional prior is integrated into the loss function for training.

4.5. Limitations of the CNN-based methods

- The *supervised method* requires large amount of ground-truth optical flow to train the parameters of the flow network to obtain reasonable accuracy, it is costly and challengingly to label the data. Besides, there is no guarantee that the trained model can still work in different scenarios with unknown challenges.
- The performance of the *unsupervised method* still has a relatively large distance compared to their supervised counterparts. Specially, the widely used surrogate losses, which depend on the fundamental assumptions of brightness constancy and spatial smoothness priors, greatly restrain the performance.
- The *semi-supervised method* requires both labeled and unlabeled data for training, thus it is easy to fall between supervised learning and unsupervised learning.

5. Conclusion

Optical flow is essential for a host of computer vision tasks. Its underspecified nature and unpractical constraints in the input introduce a number of challenges in the estimation. In this survey, we have discussed these challenges and the related methods that have been introduced to address them. We have also described the measures and datasets to benchmark optical flow estimation algorithms.

We have focused on the dominant variational algorithm and its modifications and extensions. We have outlined the main directions of research and discussed their relative advantages and disadvantages. For the first time, we have presented an in-depth overview of methods that are based on convolutional neural networks (CNNs). This novel class of approach has distinct advantages in terms of efficiency and accuracy.

The thorough understanding of the challenges in the automated estimation of optical flow and the characteristics of the described approaches will guide future research.

Acknowledgments

The work is supported by the funding CXFW-18-413100063 of Wuhan University. It is also supported by the National Natural Science Foundation of China (61501198), Natural Science Foundation of Hubei Province, China (2014CFB461) and (2017CFB598).

References

[1] B. Horn, B. Schunck, Determining optical flow, Artif. Intell. 17 (1–3) (1981) 185–203.
[2] B. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: Proc. IJCAI, 1981, pp. 674–679.
[3] 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.

[4] J. Barron, D. Fleet, S. Beauchemin, Performance of optical flow techniques, Int. J. Comput. Vis. 12 (1) (1994) 43–77.

[5] S. Baker, D. Schar, J. Lewis, S. Roth, M. Black, R. Szeliski, A database and evaluation methodology for optical flow, Int. J. Comput. Vis. 92 (1) (2011) 1–31.

[6] A. Bruhn, Variational Optic Flow Computation: Accurate Modelling and Efficient Numerics (Ph.D. thesis), Department of Mathematics and Computer Science, Saarland University, 2006.

[7] J.J. Gibson, The Perception of the Visual World, first ed., Houghton Mifflin Company, Boston, 1950.

[8] T. Poggio, W. Reiehardt, Visual control orientation behavior in the fly: Part II. Towards underlying neural interactions, Q. Rev. Biophys. 9 (1976) 377–438.

[9] I. Kajo, A. Malik, N. Kamel, An evaluation of optical flow algorithms for crowd analytics in surveillance system, in: Proc. Int. Conf. Intelligent and Advanced Systems, 2016, 2016, pp. 1–6.

[10] A. Yilmaz, O. Javed, M. Shah, Object tracking: a survey, ACM Comput. Surv. 38 (4) (2006) 1–45.

[11] C. Xu, Y. Ze, T. Shu, L. Cheng, Text detection, tracking and recognition in video: A comprehensive survey, IEEE Trans. Image Process. 25 (6) (2016) 2752–2773.

[12] F. Xiao, Y. Lee, Track and segment: An iterative unsupervised approach for video object proposals, in: Proc. CVPR, 2016, pp. 933–942.

[13] Y. Tsai, M. Yang, M. Black, Video segmentation via object flow, in: Proc. CVPR, 2016, pp. 3899–3908.

[14] Z. Tu, W. Xie, M. Yan, R.C. Veltkamp, B. Li, J. Yuan, Fusing disparate object signatures for salient object detection in video, Pattern Recognit. 72 (2017) 285–299.

[15] K. Simonyan, A. Zisserman, Two-stream convolutional networks for action recognition in videos, in: Proc. NIPS, 2014, pp. 568–576.

[16] Z. Tu, J. Cao, Y. Li, B. Li, MSR-CNN: Applying motion salient region based descriptors for action Recognition, in: Proc. ICPR, 2016, pp. 3524–3529.

[17] R. Colque, C. Caetano, M. Andrade, W. Schwartz, Histograms of optical flow orientation and magnitude and entropy to detect anomalous events in videos, IEEE Trans. Circuits Syst. Video Technol. 27 (3) (2017) 673–682.

[18] D. Xu, E. Ricci, Y. Yan, J. Song, N. Sebe, Detecting anomalous events in videos by learning deep representations of appearance and motion, Comput. Vis. Image Underst. 156 (2017) 117–127.

[19] F. Font, A. Ortiz, G. Oliver, Visual navigation for mobile robots: A survey, J. Intell. Robot. Syst. 53 (3) (2008) 263–296.

[20] G. Desouza, A. Kak, Vision for mobile robot navigation: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2) (2002) 237–267.

[21] J. Victor, G. Sandini, F. Curotto, S. Garibaldi, Divergence stereo for robot navigation: Learning from bees, in: Proc. CVPR, 1993, pp. 434–439.

[22] H. Ho, C. Wagter, B. Remes, G. de Croon, Optical flow for self-supervised learning of obstacle appearance, in: Proc. Int. Conf. Intell. Robots and Systems, 2015, pp. 3098–3104.

[23] A. Barry, High-Speed Autonomous Obstacle Avoidance with Pushbroom Stereo (Ph.D. thesis), Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2016.

[24] D. Zhou, J. Zhou, W. Fei, S. Goto, Ultra-high-throughput VLSI architecture of H.265/HEVC CABAC encoder for UHDTV applications, IEEE Trans. Circuits Syst. Video Technol. 27 (2) (2017) 380–393.

[25] K. Chen, D.A. Lorenz, Image sequence interpolation based on optical flow, segmentation, and optimal control, IEEE Trans. Image Process. 21 (3) (2012) 1020–1030.

[26] S. Niklaus, L. Mai, F. Liu, Video frame interpolation via adaptive convolution, in: Proc. CVPR, 2017, pp. 670–679.

[27] C. Liu, D. Sun, A bayesian approach to adaptive video super resolution, in: Proc. CVPR, 2011, pp. 209–216.

[28] O. Makansi, E. Ilg, T. Brox, End-to-end learning of video super-resolution with motion compensation, in: Proc. GCPR, 2017.

[29] M. Caren, S. Sandgathe, Optical flow for verification, Weather Forecast. 25 (2010) 1479–1494.

[30] P. Heas, E. Memin, 3D motion estimation of atmospheric layers from image sequences, IEEE Trans. Geosci. Remote Sens. 46 (8) (2008) 2385–2396.

[31] J. Xiong, R. Idoughi, A. Aguirre-Pablo, A. Aljedaani, X. Dun, Q. Fu, S. Thoroddsen, W. Heidrich, Rainbow particle imaging velocimetry for dense 3D fluid velocity imaging, ACM Trans. Graph. 36 (4) (2017) 36:1–14.

[32] J. Weickert, A. Bruhn, T. Brox, N. Papenberg, A survey on variational optic flow methods for small displacements, Math. Models Regist. Appl. Med. Imaging 10 (2006) 103–136.

[33] W. Trobin, Local, semi-Global, and Global Optimization for Motion Estimation (Ph.D. thesis), Institute for Computer Graphics and Vision, Graz University of Technology, Austria, 2009.

[34] Z. Tu, Variational Optical Flow Algorithms for Motion Estimation (Ph.D. thesis), Department of Information and Computing Sciences, Utrecht University, Netherlands, 2015.

[35] T. Brox, A. Bruhn, N. Papenberg, J. Weickert, High accuracy optical flow estimation based on a theory for warping, in: Proc. ECCV, 2004, pp. 25–36.

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

[37] G. Aubert, R. Deriche, P. Kornprobst, Computing optical flow via variational techniques, SIAM J. Appl. Math. 60 (1) (1999) 156–182.

[38] 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.

[39] J. Weickert, C. Schnorr, Variational optic flow computation with a spatio-temporal smoothness constraint, J. Math. Imaging Vision 14 (3) (2001) 245–255.

[40] D. Fortun, P. Bouthemy, C. Kervrann, Optical flow modeling and computation: A survey, Comput. Vis. Image Underst. 134 (2015) 1–21.

[41] Y. Boykov, O. Veksler, R. Zabih, Markov random fields with efficient approximations, in: Proc. CVPR, 1998, pp. 648–655.

[42] W. Li, D. Cosker, M. Brown, R. Tang, Optical flow estimation using Laplacian mesh energy, in: Proc. CVPR, 2013, pp. 2435–2442.

[43] M. Mozerov, Constrained optical flow estimation as a matching problem, IEEE Trans. Image Process. 22 (5) (2013) 2044–2055.

[44] M. Hornaek, F. Besse, J. Kautz, A. Fitzgibbon, C. Rother, Highly over parameterized optical flow using patchmatch belief propagation, in: Proc. ECCV, 2014, pp. 220–234.

[45] S. Roth, M. Black, On the spatial statistics of optical flow, Int. J. Comput. Vis. 74 (1) (2007) 33–50.

[46] V. Lempitsky, S. Roth, C. Rother, Fusion flow: Discrete continuous optimization for optical flow estimation, in: Proc. CVPR, 2008, pp. 1–8.

[47] M. Menze, C. Heipke, A. Geiger, Discrete optimization for optical flow, in: Proc. GCPR, 2015, pp. 16–28.

[48] A. Bruhn, J. Weickert, T. Kohlberger, C. Schnorr, A multigrid platform for real-time motion computation with discontinuity–preserving variational methods, Int. J. Comput. Vis. 70 (3) (2006) 257–277.

[49] C. Zach, T. Pock, H. Bischof, A duality based approach for realtime TV-L1 optical flow, in: DAGM conf. PR, 2007, pp. 214–223.

[50] S. Oron, A. Hillel, S. Avidan, Extended lucas-kanade tracking, in: Proc. ECCV, 2014, pp. 142–156.

[51] S. Baker, I. Matthews, Lucas-Kanade 20 years on: A unifying framework, Int. J. Comput. Vis. 56 (3) (2004) 221–255.

[52] P. Anandan, A computational framework and an algorithm for the measurement of visual motion, Int. J. Comput. Vis. 2 (3) (1989) 283–310.

[53] J. Lewis, Fast Normalized Cross-Correlation, Canadian Image Process. Pattern Recognit. Society, 1995, pp. 120–123.

[54] J. Wills, S. Belongie, A feature-based approach for determining dense long range correspondences, in: Proc. ECCV, 2004, pp. 170–182.

[55] O. Aodha, A. Humayun, M. Pollefeys, G. Brostow, Learning a confidence measure for optical flow, IEEE Trans. Pattern Anal. Mach. Intell. 35 (5) (2013) 1107–1120.

[56] J. Wills, S. Agarwal, S. Belongie, A feature-based approach for dense segmentation and estimation of large disparity motion, Int. J. Comput. Vis. 68 (2) (2006) 125–143.

[57] S. Beauchemin, J. Barron, The computation of optical flow, ACM Comput. Surv. 27 (3) (1995) 433–467.

[58] D. Heeger, Optical flow using spatiotemporal filters, Int. J. Comput. Vis. 1 (4) (1988) 279–302.

[59] A. Watson, A. Ahumada, A Look at Motion in the Frequency Domain, National Aeronautics and Space Administration, Ames Research Center, New York, 1983, pp. 1–10.

[60] E. Adelson, J. Bergen, Spatiotemporal energy models for the perception of motion, J. Opt. Soc. Amer. 2 (2) (1985) 284–299.

[61] D. Fleet, A. Jepson, Computation of component image velocity from local phase information, Int. J. Comput. Vis. 5 (1) (1990) 77–104.

[62] W. Freeman, E. Adelson, The design and use of steerable filters, IEEE Trans. Pattern Anal. Mach. Intell. 13 (9) (1991) 891–906.

[63] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazırbas, V. Golkov, P. Smagt, D. Cremers, T. Brox, FlowNet: Learning optical flow with convolutional networks, in: Proc. ICCV, 2015, pp. 2758–2766.

[64] Z. Ren, J. Yan, B. Ni, B. Liu, X. Yang, H. Zha, Unsupervised deep learning for optical flow estimation, in: Proc. AAAI, 2017.

[65] F. Guney, A. Geiger, Deep discrete flow, in: Proc. ACCV, 2016, pp. 207–224.

[66] S. Zweigand, L. Wolf, InterpoNet, a brain inspired neural network for optical flow dense interpolation, in: Proc. CVPR, 2017, pp. 4563–4572.

[67] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, T. Brox, FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks, in: Proc. CVPR, 2017, pp. 1647–1655.

[68] M. Otte, H. Nagel, Optical flow estimation: advances and comparisons, in: Proc. ECCV, 1994, pp. 51–60.

[69] B. Galvin, B. McCane, K. Novins, D. Mason, S. Mills, Recovering motion fields: An Evaluation of eight optical flow algorithms, in: Proc. BMVC, 1998, pp. 195–204.

[70] B. McCane, K. Novins, D. Crannitch, B. Galvin, On benchmarking optical flow, Comput. Vis. Image Underst. 84 (1) (2001) 126–143.

[71] R. Szeliski, Prediction error as a quality metric for motion and stereo, in: Proc. ICCV, 1999, pp. 781–788.

[72] M. Menze, A. Geiger, Object scene flow for autonomous vehicles, in: Proc. CVPR, 2015, pp. 3061–3070.

[73] D. Butler, J.W.G. Stanley, M. Black, A naturalistic open source movie for optical flow evaluation, in: Proc. ECCV, 2012, pp. 611–625.

[74] M. Aubry, D. Maturana, A. Efros, B. Russell, J. Sivic, Seeing 3D chairs: Exemplar part-based 2D–3D alignment using a large dataset of CAD models, in: Proc. CVPR, 2014, pp. 3762–3769.

[75] C. Vogel, S. Roth, K. Schindler, An evaluation of data costs for optical flow, in: Proc. GCPR, 2013, pp. 343–353.
[76] D. Sun, S. Roth, M. Black, Secrets of optical flow estimation and their principles, in: Proc. CVPR, 2010, pp. 2432–2439.
[77] L. Xu, J. Jia, Y. Matsushita, Motion detail preserving optical flow estimation, IEEE Trans. Pattern Anal. Mach. Intell. 34 (9) (2012) 1–14.
[78] J. Diaz, E. Ros, F. Pelayo, E. Ortigosa, S. Mota, FPGA-based real-time optical flow system, IEEE Trans. Circuits Syst. Video Technol. 16 (2) (2006) 274–279.
[79] L. Bao, Q. Yang, H. Jin, Fast edge-preserving patchmatch for large displacement optical flow, in: Proc. CVPR, 2014, pp. 3534–3541.
[80] L. Alvarez, R. Deriche, J. Papad, T. Sanchez, Symmetrical dense optical flow estimation with occlusions detection, Int. J. Comput. Vis. 75 (3) (2007) 371–385.
[81] A. Ayvaci, M. Raptis, S. Soatto, Sparse occlusion detection with optical flow, Int. J. Comput. Vis. 97 (3) (2012) 322–338.
[82] H. Zimmer, A. Bruhn, J. Weickert, Optic flow in harmony, Int. J. Comput. Vis. 93 (3) (2011) 368–388.
[83] Y. Mileva, A. Bruhn, J. Weickert, Illumination-robust variational optical flow with photometric invariants, in: DAGM PR Symposium, 2007, pp. 152–162.
[84] Z. Tu, R. Poppe, R.C. Veltkamp, Estimating accurate optical flow in the presence of motion blur, J. EI 24 (5) (2015).
[85] N. Papenberg, A. Bruhn, T. Brox, S. Didas, J. Weickert, Highly accurate optic flow computation with theoretically justified warping, Int. J. Comput. Vis. 67 (2) (2006) 141–158.
[86] L. Alvarez, J. Sanchez, J. Weickert, A scale-space approach to nonlocal optical flow calculations, in: Scale-Space Theories in Comput. Vis., vol. 1682, 1999, pp. 235–246.
[87] Z. Tu, W. Xie, W. Hurst, Q. Qin, Weighted root mean square approach to select the optimal smoothness parameter of the variational optical flow algorithms, Opt. Eng. 51 (3) (2012).
[88] H. Zimmer, A. Bruhn, J. Weickert, L. Valgaerts, A. Salgado, B. Rosenhahn, H. Seidel, Complementary optic flow, in: Proc. EMMCVPR, 2009, pp. 207–220.
[89] J. Weijer, T. Gevers, Robust optical flow from photometric invariants, in: Proc. ICIP, 2004, pp. 1835–1838.
[90] M. Mohamed, H. Rashwan, B. Mertsching, M. Garcia, D. Puig, Illumination-Robust optical flow using a local directional pattern, IEEE Trans. Circuits Syst. Video Technol. 24 (9) (2014) 1499–1508.
[91] J. Aujol, G. Gilboa, T. Chan, S. Osher, Structure-texture image decomposition-modeling, algorithms, and parameter selection, Int. J. Comput. Vis. 67 (1) (2006) 111–136.
[92] A. Wedel, T. Pock, C. Zach, D. Cremers, H. Bischof, An improved algorithm for tv-l1 optical flow, Sta. and Geometrical Appl. to Vis. Motion Anal. 5064 (2008) 23–45.
[93] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D 60 (1992) 259–268.
[94] P. Golland, A. Bruckstein, Motion from color, Comput. Vis. Image Underst. 68 (3) (1997) 346–362.
[95] M.J. Black, P. Anandan, Robust dynamic motion estimation over time, in: Proc. CVPR, 1991, pp. 292–302.
[96] P. Charbonnier, L. Blanc-Feraud, G. Aubert, M. Barlaud, Two deterministic half-quadratic regularization algorithms for computed imaging, in: Proc. ICIP, 1994, pp. 168–172.
[97] J. Odobez, P. Bouthemy, Robust multiresolution estimation of parametric motion models, J. Vis. Commun. Image Represent. 6 (4) (1995) 348–365.
[98] E. Memin, P. Perez, Dense estimation and object-based segmentation of the optical flow with robust techniques, IEEE Trans. Image Process. 7 (5) (1998) 703–719.
[99] T. Senst, V. Eiselein, T. Sikora, Robust local optical flow for feature tracking, IEEE Trans. Circuits Syst. Video Technol. 22 (9) (2012) 1377–1387.
[100] D. Sun, S. Roth, M. Black, A quantitative analysis of current practices in optical flow estimation and the principles behind them, Int. J. Comput. Vis. 106 (2) (2014) 115–137.
[101] A. Blake, A. Zisserman, Visual Reconstruction, The MIT Press, Cambridge, MA, 1987.
[102] J. Weickert, C. Schnorr, A theoretical framework for convex regularizers in PDE-based computation of image motion, Int. J. Comput. Vis. 45 (3) (2001) 245–264.
[103] L. Alvarez, J. Weickert, J. Sanchez, Reliable estimation of dense optical flow fields with large displacements, Int. J. Comput. Vis. 39 (1) (2000) 41–56.
[104] C. Schnorr, Segmentation of visual motion by minimizing convex non-quadratic functionals, in: Proc. ICPR, 1994, pp. 661–663.
[105] M. Werlberger, T. Pock, H. Bischof, Motion estimation with non-local total variation regularization, in: Proc. ICCV, 2010, pp. 2464–2471.
[106] P. Krahenbuhl, V. Koltun, Efficient nonlocal regularization for optical flow, in: Proc. ECCV, 2012, pp. 356–369.
[107] S. Volz, A. Bruhn, L. Valgaerts, H. Zimmer, Modeling temporal coherence for optical flow, in: Proc. ICCV, 2011, pp. 1116–1123.
[108] R. Garg, A. Roussos, L. Agapito, Robust trajectory-space TV-L1 optical flow for non-rigid sequences, in: Proc. EMMCVPR, 2011, pp. 300–314.
[109] 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.
[110] V. Solo, A sure-fired way to choose smoothing parameters in ill-conditioned inverse problems, in: Proc. ICIP, 1996, pp. 89–92.
[111] L. Ng, V. Solo, A data-driven method for choosing smoothing parameters in optical flow problems, in: Proc. ICIP, 1997, pp. 360–363.
[112] L. Raket, Local smoothness for global optical flow, in: Proc. ICIP, 2012, pp. 1–4.
[113] Z. Tu, R. Poppe, R.C. Veltkamp, Weighted local intensity fusion method for variational optical flow estimation, Pattern Recognit. 50 (2016) 223–232.
[114] D. Sun, E. Sudderth, H. Pfister, Layered RGBD scene flow estimation, in: Proc. CVPR, 2015, pp. 548–556.
[115] J. Rua, T. Crivelli, P. Bouthemy, P. Perez, Determining occlusions from space and time image reconstructions, in: Proc. CVPR, 2016, pp. 1382–1391.
[116] V. Estellers, S. Soatto, Detecting occlusions as an inverse problem, J. Math. Imaging Vision 54 (2) (2015) 181–198.
[117] J. Xiao, H. Cheng, H. Sawhney, C. Rao, M. Isnardi, Bilateral filtering–based optical flow estimation with occlusion detection, in: Proc. ECCV, 2006, pp. 211–224.
[118] D. Sun, S. Roth, J. Lewis, J. Black, Learning optical flow, in: Proc. ECCV, 2008, pp. 83–97.
[119] C. Zitnick, T. Kanade, A cooperative algorithm for stereo matching and occlusion detection, IEEE Trans. Pattern Anal. Mach. Intell. 22 (7) (2000) 675–684.
[120] S. Ince, J. Konrad, Occlusion-aware optical flow estimation, IEEE Trans. Image Process. 17 (8) (2008) 1443–1451.
[121] P. Sand, S. Teller, Particle video: Long-range motion estimation using point trajectories, Int. J. Comput. Vis. 80 (1) (2008) 72–91.
[122] P. Smith, T. Drummond, R. Cipolla, Layered motion segmentation and depth ordering by tracking edges, IEEE Trans. Pattern Anal. Mach. Intell. 26 (4) (2004) 479–493.
[123] A. Stein, M. Hebert, Occlusion boundaries from motion: Low-level detection and mid-level reasoning, Int. J. Comput. Vis. 82 (3) (2009) 325–357.
[124] P. Sundberg, T. Brox, M. Maire, P. Arbelaez, J. Malik, Occlusion boundary detection and figure/ground assignment from optical flow, in: Proc. CVPR, 2011, pp. 2233–2240.
[125] A. Humayun, O. Aodha, G. Brostow, Learning to find occlusion regions, in: Proc. CVPR, 2011, pp. 2161–2168.
[126] C. Kondermann, R. Mester, C. Garbe, A statistical confidence measure for optical flows, in: Proc. ECCV, 2008, pp. 290–301.
[127] F. Xu, Q. Dai, Occlusion-aware motion layer extraction under large interframe motions, IEEE Trans. Image Process. 20 (9) (2011) 2615–2626.
[128] D. Sun, E. Sudderth, M. Black, Layered segmentation and optical flow estimation over time, in: Proc. CVPR, 2012, pp. 1768–1775.
[129] E. Lobaton, R. Vasudevan, R. Bajcsy, R. Alterovitz, Local occlusion detection under deformations using topological invariants, in: Proc. ECCV, 2010, pp. 101–114.
[130] W. Thompson, Exploiting discontinuities in optical flow, Int. J. Comput. Vis. 30 (3) (1998) 163–173.
[131] C. Zhang, Z. Chen, M. Wang, M. Li, S. Jiang, Robust non-local TV-L1 optical flow estimation with occlusion detection, IEEE Trans. Image Process. 26 (8) (2017) 4055–4066.
[132] X. Shen, Y. Wu, Sparsity model for robust optical flow estimation at motion discontinuities, in: Proc. CVPR, 2010, pp. 2456–2463.
[133] K. Jia, X. Wang, X. Tang, Optical flow estimation using learned sparse model, in: Proc. ICCV, 2011, pp. 2391–2398.
[134] E. Memin, P. Perez, Hierarchical estimation and segmentation of dense motion fields, Int. J. Comput. Vis. 46 (2) (2002) 129–155.
[135] T. Amiaz, N. Kiryati, Piecewise-smooth dense optical flow via level sets, Int. J. Comput. Vis. 68 (2) (2006) 111–124.
[136] L. Xu, J. Chen, J. Jia, A segmentation based variational model for accurate optical flow estimation, in: Proc. ECCV, 2008, pp. 671–684.
[137] L. Lara, D. Sun, V. Jampani, M. Black, Optical flow with semantic segmentation and localized layers, in: Proc. CVPR, 2016, pp. 3889–3898.
[138] F. Steinbrucker, T. Pock, Large displacement optical flow computation without warping, in: Proc. CVPR, 2009, pp. 1069–1074.
[139] Z. Chen, H. Jin, Z. Lin, S. Cohen, Y. Wu, Large displacement optical flow from nearest neighbor fields, in: Proc. CVPR, 2013, pp. 2443–2450.
[140] Z. Tu, N. van der Aa, C. van Gemeren, R.C. Veltkamp, A combined post-filtering method to improve accuracy of variational optical flow estimation, Pattern Recognit. 47 (5) (2014) 1926–1940.
[141] P. Anandan, Measuring Vision Motion from Image Sequence (Ph.D. thesis), University of Massachusetts, US, 1987.
[142] A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger, C. Schnorr, Variational optical flow computation in real time, IEEE Trans. Image Process. 14 (5) (2005) 608–615.
[143] Y. Yang, S. Soatto, S2F: Slow-To-Fast Interpolator Flow, in: Proc. CVPR, 2017, pp. 2087–2096.
[144] C. Liu, J. Yuen, A. Torralba, SIFT flow: Dense correspondence across different scenes and its applications, IEEE Trans. Pattern Anal. Mach. Intell. 33 (5) (2011) 978–994.
[145] D. Lowe, Object recognition from local scale-invariant features, in: Proc. ICCV, 1999, pp. 1150–1157.
[146] P. Weinzaepfel, J. Revaud, Z. Harchaoui, C. Schmid, DeepFlow: Large displacement optical flow with deep matching, in: Proc. ICCV, 2013, pp. 1385–1392.
[147] P. Heas, E. Memin, N. Papadakis, A. Szantai, Layered estimation of atmospheric mesoscale dynamics from satellite imagery, IEEE Trans. Geosci. Remote Sens. 45 (12) (2007) 4087–4104.
[148] M. Stoll, S. Volz, A. Bruhn, Adaptive integration of feature matches into variational optical flow methods, in: Proc. ACCV, 2012, pp. 1–14.

[149] J. Zin, R. Dupont, A. Bartoli, A general dense image matching framework combining direct and feature-based costs, in: Proc. ICCV, 2013, pp. 185–192.
[150] Z. Wang, F. Wu, Z. Hu, MSLD: A robust descriptor for line matching, Pattern Recognit 42 (5) (2009) 941–953.
[151] J. Revaud, P. Weinzaepfel, Z. Harchaoui, C. Schmid, DeepMatching: hierarchical deformable dense matching, Int. J. Comput. Vis. 120 (3) (2016) 300–323.
[152] T. Kroeger, R. Timofte, D. Dai, L. Gool, Fast optical flow using dense inverse search, in: Proc. ECCV, 2016, pp. 471–488.
[153] C. Barnes, E. Shechtman, D. Goldman, A. Finkelstein, The generalized patchmatch correspondence algorithm, in: Proc. ECCV, 2010, pp. 29–43.
[154] J. Revaud, P. Weinzaepfel, Z. Harchaoui, C. Schmid, EpicFlow: Edge-preserving interpolation of correspondences for optical flow, in: Proc. CVPR, 2015, pp. 1164–1172.
[155] Y. Hu, Y. Li, R. Song, Robust interpolation of correspondences for large displacement optical flow, in: Proc. CVPR, 2017, pp. 481–489.
[156] Y. Li, Pyramidal gradient matching for optical flow estimation, arXiv preprint, 2017, arXiv:1704.03217.
[157] J. Lu, Y. Li, H. Yang, D. Min, W. Eng, M. Do, PatchMatch filter: Edge-aware filtering meets randomized search for visual correspondence, IEEE Trans. Pattern Anal. Mach. Intell. 39 (9) (2016) 1866–1879.
[158] C. Bailer, B. Taetz, D. Stricker, Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation, in: Proc. ICCV, 2015, pp. 4015–4023.
[159] C. Barnes, E. Shechtman, A. Finkelstein, D. Goldman, PatchMatch: A randomized correspondence algorithm for structural image editing, ACM Trans. on Graphics 28 (3) (2009) 24:1–11.
[160] Y. Hu, R. Song, Y. Li, Efficient coarse-to-fine patchmatch for large displacement optical flow, in: Proc. CVPR, 2016, pp. 5704–5712.
[161] P. Hammer, P. Hansen, B. Simeone, Roof duality, complementation and persistency in quadratic 0-1 optimization, Math. Program. 28 (1984) 121–155.
[162] C. Rother, V. Kolmogorov, V. Lempitsky, M. Szummer, Optimizing binary MRFs via extended roof duality, in: Proc. CVPR, 2007, pp. 1–8.
[163] C. Fermuller, D. Shulman, Y. Aloimonos, The statistics of optical flow, Comput. Vis. Image Underst. 82 (1) (2001) 1–32.
[164] A. Hadiashar, R. Tennakoon, M. Bruijne, Quantification of Smoothing Requirement for 3D optic flow calculation of volumetric images, IEEE Trans. Image Process. 22 (6) (2013) 2128–2137.
[165] J. Weber, J. Malik, Robust computation of optical flow in a multi-scale differential framework, Int. J. Comput. Vis. 14 (1) (1995) 67–81.
[166] 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.
[167] X. Song, L.D. Seneviratne, K. Althoefer, A Kalman filter-integrated optical flow method for velocity sensing of mobile robots, IEEE Trans. Mechatronics 16 (3) (2011) 551–563.
[168] K. He, J. Sun, X. Tang, Guided image filtering, IEEE Trans. Pattern Anal. Mach. Intell. 35 (6) (2013) 1397–1409.
[169] Z. Tu, R. Poppe, R.C. Veltkamp, Adaptive guided image filter for warping in variational optical flow, Signal Proc. 127 (2016) 253–265.
[170] A. Buades, B. Coll, J. Morel, A non-local algorithm for image denoising, in: Proc. CVPR, 2005, pp. 60–65.
[171] C. Liu, W. Freeman, A high-quality video denoising algorithm based on reliable motion estimation, in: Proc. ECCV, 2010, pp. 706–719.
[172] A. Wedel, D. Cremers, T. Pock, H. Bischof, Structure- and motion-adaptive regularization for high accuracy optic flow, in: Proc. ICCV, 2009, pp. 1663–1668.
[173] Z. Tu, C. van Gemeren, R.C. Veltkamp, Improved color patch similarity measure based weighted median filter, in: Proc. ACCV, 2015, pp. 1–15.
[174] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: Proc. ICCV, 1998, pp. 839–846.
[175] M. Elad, A. Feuer, Recursive optical flow estimation-adaptive filtering approach, J. Vis. Commun. Image Represent. 9 (2) (1998) 119–138.
[176] C. Rabe, T. Muller, A. Wedel, U. Franke, Dense, robust and accurate motion field estimation from stereo image sequences in real-time, in: Proc. ECCV, 2010, pp. 582–595.
[177] T. Portz, L. Zhang, H. Jiang, Optical flow in the presence of spatially-varying motion blur, in: Proc. CVPR, 2012, pp. 1752–1759.
[178] Z. Tu, W. Xie, J. Cao, R. Poppe, R.C. Veltkamp, Variational method for joint optical flow estimation and edge-aware image restoration, Pattern Recognit. 65 (2017) 11–25.
[179] J. Xu, R. Ranftl, V. Koltun, Accurate optical flow via direct cost volume processing, in: Proc. CVPR, 2017, pp. 1289–1297.
[180] W. Trobin, T. Pock, D. Cremers, H. Bischof, An unbiased second-order prior for high-accuracy motion estimation, in: DAGM Symposium on PR, 2008, pp. 396–405.
[181] 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: Proc. CVPR, 2011, pp. 1193–1200.
[182] R. Palomares, E. Llopis, C. Ballester, A new minimization strategy for large displacement variational optical flow, J. Math. Imaging Vision 58 (2017) 27–46.
[183] Y. Niu, OPtical Flow Estimation in the Presence of Fast or Discontinuous Motion, School of Computer Science (Ph.D. thesis), University of Adelaide, Australia, 2010.
[184] G. Aubert, P. Kornprobst, Mathematical problems in image processing: Partial differential equations and the calculus of variations, in: Applied mathematical sciences, Springer–Verlag New York, 2006.
[185] H. Ho, R. Goecke, Optical flow estimation using Fourier Mellin Transform, in: Proc. CVPR, 2008, pp. 1–8.
[186] Y. Niu, A. Dick, M. Brooks, Locally oriented optical flow computation, IEEE Trans. Image Process. 21 (4) (2012) 1573–1586.
[187] W. Qiu, X. Wang, X. Bai, A. Yuille, Z. Tu, Scale-space SIFT flow, in: Proc. WACV, 2014, pp. 1112–1119.
[188] A. Baghaie, R. Souza, Z. Yu, Dense descriptors for optical flow estimation: A comparative study, J. Imaging 3 (1) (2017) 1–19.
[189] S. Meister, J. Hur, S. Roth, Unflow: unsupervised learning of optical flow with a bidirectional census loss, in: Proc. AAAI, 2018.
[190] E. Ilg, O. Cicek, S. Galesso, A. Klein, O. Makansi, F. Hutter, T. Brox, Uncertainty estimates and multi-hypotheses networks for optical flow, in: Proc. ECCV, 2018, pp. 652–667.
[191] N. Mayer, P.H. E. Ilg, P. Fischer, D. Cremers, A. Dosovitskiy, T. Brox, A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation, in: Proc. CVPR, 2016, pp. 4040–4048.
[192] X. Xiang, M. Zhai, R. Zhang, Y. Qiao, A.E. Saddi, Deep optical flow supervised learning with prior assumptions, IEEE Access 6 (2018) 43222–43232.
[193] A. Ranjan, M. Black, Optical flow estimation using a spatial pyramid network, in: Proc. CVPR, 2017, pp. 4161–4170.
[194] D. Sun, X. Yang, M. Liu, J. Kautz, PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume, in: Proc. CVPR, 2018, pp. 8934–8943.
[195] T. Hui, X. Tang, C. Loy, Liteflownet: A lightweight convolutional neural network for optical flow estimation, in: Proc. CVPR, 2018, pp. 8981–8989.
[196] S. Zhao, L. X, O. Bourahla, Deep optical flow estimation via multi-scale correspondence structure learning, in: Proc. IJCAI, 2017.
[197] D. Gadot, L. Wolf, PatchBatch: A batch augmented loss for optical flow, in: Proc. CVPR, 2016, pp. 4236–4245.
[198] T. Schuster, L. Wolf, D. Gadot, Optical flow requires multiple strategies (but only one network), in: Proc. CVPR, 2017, pp. 4950–4959.
[199] M. Bai, W. Luo, K. Kundu, R. Urtasun, Exploiting semantic information and deep matching for optical flow, in: Proc. ECCV, 2016, pp. 154–170.
[200] C. Bailer, K. Varanasi, D. Stricker, CNN-based patch matching for optical flow with thresholded hinge embedding loss, in: Proc. CVPR, 2017, pp. 3250–3259.
[201] P. Hu, G. Wang, Y.P. Tan, Recurrent spatial pyramid cnn for optical flow estimation, IEEE Trans. Multimedia 20 (10) (2018) 2814–2823.
[202] A. Ahmadi, I. Patras, Unsupervised convolutional neural networks for motion estimation, in: Proc. ICIP, 2016, pp. 1629–1633.
[203] J.J. Yu, A.W. Harley, K.G. Derpanis, Back to basics: unsupervised learning of optical flow via brightness constancy and motion smoothness, in: Proc. ECCVW, 2016, pp. 3–10.
[204] Y. Zhu, Z. Lan, S. Newsamy, A. Hauptmann, Guided optical flow learning, in: Proc. CVPRW, 2017.
[205] M. Jaderberg, K. Simonyan, A. Zisserman, K. Kavukcuoglu, Spatial transformer networks, in: Proc. NIPS, 2015, pp. 2017–2025.
[206] L. Fan, W. Huang, C. Gan, S. Ermon, B. Gong, J. Huang, End-to-end learning of motion representation for video understanding, in: Proc. CVPR, 2018, pp. 6016–6025.
[207] Y. Wang, Y. Yang, Z. Yang, L. Zhao, P. Wang, W. Xu, Occlusion aware unsupervised learning of optical flow, in: Proc. CVPR, 2018, pp. 4884–4893.
[208] M. Caron, P. Bojanowski, A. Joulin, M. Douze, Deep clustering for unsupervised learning of visual features, in: Proc. ECCV, 2018, pp. 132–149.
[209] W. Lai, J. Huang, M. Yang, Semi-supervised learning for optical flow with generative adversarial networks, in: Proc. NIPS, 2017.
[210] Y. Yang, S. Soatto, Conditional prior networks for optical flow, in: Proc. ECCV, 2018, pp. 271–287.

Zhigang Tu started his Master Degree in image processing at the School of Electronic Information, Wuhan University, China, 2008. In 2015, he received the Ph.D. degree in Computer Science from Utrecht University, Netherlands. From 2015 to 2016, he was a postdoctoral researcher at Arizona State University, US. Then from 2016 to 2018, he was a research fellow at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He is currently a professor at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote sensing, Wuhan University. His research interests include computer vision, image processing, video analytics, and machine learning. Specially for motion estimation, object segmentation, ~~object tracking~~, action recognition and localization, and anomaly detection.

Wei Xie is an associate professor at Computer School of Central China Normal University, China. His research interests include motion estimation, superresolution reconstruction, image fusion and image enhancement. He received his BE degree in electronic information engineering and PhD degree in communication and information system from Wuhan University, China, in 2004 and 2010, respectively. Then, from 2010 to 2013, he served as an assistant professor at Computer School of Wuhan University, China.

Dejun Zhang received the bachelor in communication engineering at the School of Information Science and Technology, Southwest Jiaotong University, China, 2006. In 2011, he received the Master degree in electronic engineering at the School of Manufacturing Science and Engineering, Southwest University of Science and Technology, China. In 2015, he received the Ph.D. degree in Computer Science from Wuhan University, China. He is currently a lecturer with the faculty of college of information and engineering, Sichuan agricultural university, China. Since 2015, he has been serving as a senior member of the China Society for Industrial and Applied Mathematics (CSIAM) and a committee member of the geometric design and computing of CSIAM. His research areas include machine learning, bioinformatics and computer graphics. His research interests include digital geometric processing, computer graphic, action recognition and localization.

Ronald Poppe received a Ph.D. in Computer Science from the University of Twente, The Netherlands. He was a visiting researcher at the Delft University of Technology, Stanford University and University of Lancaster. He is currently an assistant professor at the Information and Computing Sciences department of Utrecht University. His research interests include the analysis of human behavior from videos and other sensors, the understanding and modeling of human (communicative) behavior and the applications of both in real-life settings. In 2012 and 2013, he received the most cited paper award from the ‘‘Image and Vision Computing" journal, published by Elsevier.

Remco C. Veltkamp is full professor of Multimedia at Utrecht University, Netherlands. His research interests are the analysis, recognition and retrieval of, and interaction with, music, images, and 3D objects and scenes, in particular the algorithm and experimentation aspects. He has written over 150 refereed papers in reviewed journals and conferences, and supervised 15 PhD theses. He was director of the national project GATE — Game Research for Training and Entertainment.

Baoxin Li received the Ph.D. degree in electrical engineering from the University of Maryland, College Park, in 2000. He is currently a full professor and Chair of computer science and engineering with Arizona State University, US. From 2000 to 2004, he was a senior researcher with SHARP Laboratories of America, Camas, WA, where he was the technical lead in developing SHARP’s HiIMPACT Sports technologies. From 2003 to 2004, he was also an adjunct professor with the Portland State University, Portland, OR. He holds nine issued US patents. His current research interests include computer vision and pattern recognition, image/video processing, multimedia, medical image processing, and statistical methods in visual computing. He won the SHARP Laboratories’ President Award twice, in 2001 and 2004. He also received the SHARP Laboratories’ Inventor of the Year Award in 2002. He received the National Science Foundation’s CAREER Award from 2008 to 2009. He is a senior member of the IEEE.

Junsong Yuan (M’08–SM’14) received his Ph.D. from Northwestern University and M.Eng. from National University of Singapore. Before that, he graduated from the Special Class for the Gifted Young of Huazhong University of Science and Technology, Wuhan, China, in 2002. He is currently an associate professor at Computer Science and Engineering department of State University of New York at Buffalo. Previously, he was an associate professor at Nanyang Technological University (NTU), Singapore. His research interests include computer vision, video analytics, gesture and action analysis, large-scale visual search and mining. He received best paper awards from ICAR 2017, IEEE Transactions on Multimedia 2016, a Doctoral Spotlight Award from CVPR 2009 and the Outstanding EECS Ph.D. Thesis award from Northwestern University. He is currently Senior Area Editor of Journal of Visual Communications and Image Representations (JVCI), Associate Editor of IEEE Transactions on Image Processing (T-IP), IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT), and served as Guest Editor of International Journal of Computer Vision (IJCV).