Review Paper

Traditional and modern strategies for optical flow: an investigation

Syed Tafseer Haider Shah[1] · Xiang Xuezhi[1]

Received: 14 March 2020 / Accepted: 15 January 2021 / Published online: 8 February 2021

Abstract
Optical Flow Estimation is an essential component for many image processing techniques. This field of research in computer vision has seen an amazing development in recent years. In particular, the introduction of Convolutional Neural Networks for optical flow estimation has shifted the paradigm of research from the classical traditional approach to deep learning side. At present, state of the art techniques for optical flow are based on convolutional neural networks and almost all top performing methods incorporate deep learning architectures in their schemes. This paper presents a brief analysis of optical flow estimation techniques and highlights most recent developments in this field. A comparison of the majority of pertinent traditional and deep learning methodologies has been undertaken resulting the detailed establishment of the respective advantages and disadvantages of the traditional and deep learning categories. An insight is provided into the significant factors that affect the success or failure of the two classes of optical flow estimation. In establishing the foremost existing and inherent challenges with traditional and deep learning schemes, probable solutions have been proposed indeed.

Keywords Optical flow · Variational methods · Convolutional neural networks · Deep learning

1 Introduction

A fundamental component in the construction of a machine's vision system is the computation of optical flow which is obtained by estimating a dense motion field corresponding to the displacement of each pixel in consecutive frames of an image sequence. Its reliable calculation comprises one of the main challenges in computer vision. Optical flow can be combined with various computer vision tasks such as video coding, segmentation, tracking [1] and multi view-reconstruction [2]. Some other fields where optical flow has played an important role includes fluid mechanics [3], solar physics [4], autonomous driving [5], biomedical images [6], breast tumors [7], bladder cancer [8] surveillance and traffic monitoring [9], virtual reality [10], face recognition and tracking [11], and action recognition videos [12].

Optical flow is the pattern of the apparent motion of objects in a visual scene caused by the motion of an object or camera or both. When a camera records a scene for a given time, the resulting image sequence can be considered as a function of gray values at image pixel position (x, y) and the time t. If the camera or an object moves within the scene, this motion results in a time-dependent displacement of the gray values in the image sequence. The resulting two-dimensional apparent motion field in the image domain is the Optical Flow Field. Figure 1 shows an image sequence and the corresponding optical flow field in color and the arrow visualizations.

At present, optical flow estimation stands at its peak with a steady progress. In last 4 decades, a whole class of various techniques and novel concepts has evolved in this area. Particularly, remarkable development has been witnessed in the last decade. On one hand, the advance level datasets such as Middlebury [13], MPI-Sintel [14]and

✉ Syed Tafseer Haider Shah, Tesla1835@gmail.com | [1]College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China.

SN Applied Sciences (2021) 3:289 | https://doi.org/10.1007/s42452-021-04227-x

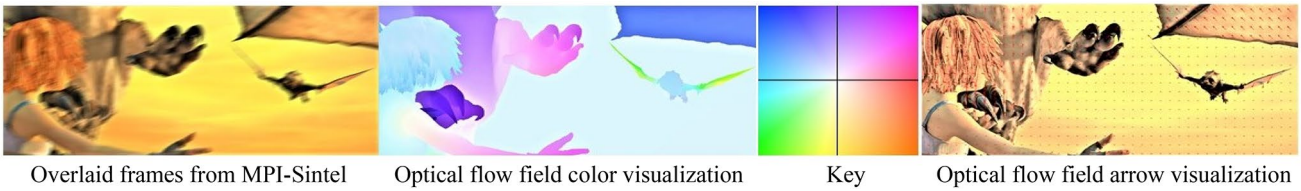


Fig. 1 Overlaid frames from MPI-Sintel and the corresponding optical flow field

KITTI [15, 16] presented substantial novel challenges for the optical flow algorithms, on the other hand traditional methods such as LDOF [17], DeepFlow [18], EpicFlow [19], DiscreteFlow [20], FlowFields [21] and MirrorFlow [22] came up with a significant number of novel strategies. These innovations solved the correspondence problem with outstanding performance. However, the enhanced accuracy subsequently increased the evaluation time. Consequently, none of the major traditional methods runs in real time currently. The high computational cost of modern traditional techniques is a major limitation for their adoption into other applications. In being realistic, there was a requirement of a unique and extraordinarily different approach than the contemporary methods to be incorporated. Meanwhile, easy access to computationally powerful Graphic Processing Units (GPUs) drew researchers' attention to focus on deep learning side. FlowNet [23] presented the first ever deep learning scheme based on convolutional neural networks. The scheme seems promising as it runs in real time. However, its efficiency remained below the state of art traditional methods. FlowNet2 [24] built a stronger structure by stacking together multiple FlowNet [23] modules which out-performed many traditional methods such as DeepFlow [18] and EpicFlow [19] on MPI-Sintel benchmark [14]. However, the large model-design of FlowNet2 requires heavy memory; hence is not well-suited for mobile and other embedded devices. Later work focused on designing light-weight modules without compromising on accuracy of the output. This was achieved through borrowing many popular ideas from traditional methods and implanting into deep learning schemes, stimulating a great amount of work. SpyNet [25] combines coarse-to-fine approach of traditional methods and deep learning methods, LiteFlow [26] uses traditional brightness inconsistency map to tackle occlusions, PWC-Net [27] integrates traditional stereo matching, feature extraction and cost volume with deep learning, Continual Flow [28] combines occlusions and cost volume together with optical flow, MFF [29] fuses the warped optical flow from previous frame with the current optical flow estimate, SelFlow [30] uses flow estimation from non-occluded pixels for self-learning, IRR [31] learns to refine its previous estimate by iteratively re-using the same network block with shared weights. At present, optical flow schemes based on deep learning have completely outperformed the traditional methods in accuracy and run time. In fact, deep learning methods run in real time with much higher accuracy. The extraordinary performance of deep learning methods is due to various ideas and novel concepts already adopted by traditional schemes.

A brief analysis of classical and modern techniques is presented here. It also provides a comparison of the benefits and limitations of different types, with a discussion of major inherent and potential challenges and their possible solutions. In the paper Sect. 2 describes methodology, different approaches and popular techniques for optical flow estimation. Section 3 gives details on evaluation benchmarks, assessment parameters, and commonly used datasets in this field. Section 4 is about major challenges for optical flow, and also highlights the solutions proposed by mainstream methods. Section 5 provides our outlook about modern challenges with an end-note for future work. Section 6 concludes this paper.

2 Methodology

At present, there are two major approaches for estimating optical flow. First is traditional approach that implies handcraft feature evaluation schemes into the main framework; second, the convolutional neural network approach based on deep learning principles.

2.1 Traditional methods

These methods dominated the field of optical flow for almost 4 decades. Traditional methods.

can further be divided into the following classes: pixel based [20, 21, 32] feature based [17–19] and energy based classes [33–35], however, there are no explicit limits to separate one class from another. Among them, the most successful and widely used techniques are variational methods. These schemes estimate optical flow by minimizing an energy functional derived on the basis of brightness constancy and smoothness

assumptions [33]. Variational methods can further be divided into global and local categories.

2.1.1 Global techniques

Are based on brightness constancy and smoothness assumptions; this approach builds an energy functional whose minimizing scheme yields the flow field. In 1981 Horn and Schunck [33] solved the under determined aperture problem by providing additional smoothness constraint. Horn and Schunck's approach was followed by many researchers; however, the algorithm faces difficulties in many practical situations such as varying illumination and large displacements.

2.1.2 Local (total least square) techniques

These approaches take the assumption of constant essential flow within a small neighborhood "R" consisting of "n x n" pixels. The flow constraint is evaluated at all pixels within the neighborhood window and the resulting equations are solved using the least square method. The objective energy to be minimized is the weighted sum of the potentials provided by each pixel of R.

2.1.3 Local versus global

The local approaches have many advantages. Firstly, the search for the flow vectors gives a good estimate without considering the entire image. Secondly, global methods produce erroneous results for non-homogeneous regions because the flow vectors evaluated by the iterative schemes of global methods tend to spread out. For instance, two occluding objects passing, both having different orthogonal components but similar spatial gradients will be merged by a global method at the boundary, consequently losing the sharp discontinuity. The vectors produced by a local method will not go through this problem.

2.1.4 Coarse to fine or pyramid schemes

The linearization adopted by differential schemes requires small displacements. For large motions this constraint is violated and produce errors. The standard practice to deal with this issue is to carry out the estimation process in a coarse-to-fine framework [34]. A pyramid of coarse-to-fine down-sampled versions of the original image is created by filtering and re-sampling the images at lower resolution. A coarse match is done at the lower level which is used to define a small area with the next higher resolution. The solution is iteratively refined until reaching the full image resolution. Most modern methods choose this strategy for large displacements [19, 25, 36]. Though, earlier researchers had adopted this technique on empirical grounds, the most prominent work was carried out by Thomas Brox [34] who provided a theoretical ground to integrate the variational methods with coarse-to-fine.

2.1.5 Limitations of coarse to fine schemes

The coarse to fine strategies enormously improved the performance. However, they do possess intrinsic limitations. For instance, they may lead to a solution trapped into local minima. Secondly the objects, whose extents are smaller than their respective displacements, may be lost at coarser levels due to smoothing process (Fig. 2). A third weakness is error-propagation. At coarser levels, different motion layers can overlap, and may propagate across scales. A prominent alternative to coarse to fine schemes is discrete optimization [20] as is adopted by many stereo matching methods. However, optical flow requires full data cost volume while stereo matching does not require the image pyramid scheme because it is 1D problem. The optical flow estimation being 2D involves extremely large size of the label space, making the estimation process difficult with discrete optimization.

2.1.6 Feature based methods

Although variational methods are among the most popular techniques, accuracy of the vectors generated by these methods is always uncertain. The only region that

(a)Original frame (b) Optical flow computed with a warping method (c) Ground truth

Fig. 2 Failure of the pyramid schemes for fast motion. The fast moving arm has disappeared in the evaluated flow field in (**b**)

can produce dependable flow vectors for a unique correspondence is one consisting of points with enough gradients in two directions. In a plane, uniform and homogeneous region the flow vectors are undefined and may give erroneous results. These observations stimulated researchers to ignore all unreliable regions, keeping only trustworthy vectors; giving birth to a new class of optical flow methods known as feature based methods [17–19].

2.1.7 Advantages and limitations of feature based methods

Feature based methods helped to overcome the major problem of large displacements [17]. Feature matching is similar to the local parametric approach, although the main difference lies in the optimization process. The linearization formulation involves differential optimization while feature matching walks around a discrete space of correspondence. Feature matching can handle large displacements without adopting pyramid schemes which is an added benefit. However, they do have limitations that include excessive computational cost on the searching process [19, 22, 37], large errors induced by repetitive textures, and reduced accuracy because of the integer displacements [20] due to possibly sparse set of correspondences. For the estimation of a dense motion field, feature matching can be divided into two major groups:

2.1.8 Feature matching for local filtering

A key problem for local filtering is the excessive computational time in search space. Numerous attempts to overcome this include: search in trees, multi-scale search and integral images. A milestone in this field is the introduction of Patch Match [38]. Primarily designed for image editing and later applied to other applications; Patch match yielded inspiring results with respect to the accuracy and low computation cost.

2.1.9 Feature matching for global framework

When global methods are combined with coarse-to-fine schemes, the models face a major problem of losing small and fast moving objects due to smoothing process at coarser levels (Fig. 2). LDOF [17], addressed this long existing problem by adding a new constraint, and showed that feature matching can be integrated with global frameworks to overcome this smoothing out problem. In the follow up work many researchers [18, 21, 39] utilized the LDOF approach.

2.2 Deep learning based or CNN methods

Based on machine learning principles, these algorithms learn to compute optical flow from a pair of input images. In recent years convolutional networks have been used to estimate the optical flow with promising results [23–28, 30]. Convolutional neural nets can go through supervised or unsupervised training for per-pixel image classification. These tasks are similar to optical flow estimation in the sense that they involve per-pixel predictions. However, optical flow estimation requires the network to learn feature representations and match them at different locations in two images. In this sense it is different from the previous applications of convolutional neural networks.

With respect to functioning, convolutional neural nets equipped with multi-layers can extract intangible and multi-scale features. The main disadvantage of deep learning method is their requirement of large quantity of labelled training data. Until now, researchers have been relying on a synthetically rendered dataset but these datasets do not reflect the genuine photometric properties of real video sequences which is a major challenge for deep learning methods. Another key disadvantage of deep learning methods is their necessity for a large number of parameters. On one hand this results in huge memory footprint, on the other hand it causes over fitting. The over excessive memory, and millions of parameters can adversely affect the network's performance and learning of the algorithm.

2.2.1 Supervised and unsupervised learning

Supervised methods comprise a major class of the deep learning category that gives better performance in terms of accuracy and run time [23–27]. Supervised learning for optical flow requires labeling for training algorithms. Besides being tedious, a major challenge to this approach is the non-availability of real-world datasets annotated with ground truth, large enough to train a consistent model [24]. Existing datasets [13–15] are too small to support training. Computer generated synthetic scenes and their corresponding ground truth have been used by [23, 24]. Creating such big-sized diverse imagery is not only expensive and laborious, but the algorithms trained on synthetic data will not be successful when it comes across real world photometric effects such as illumination variations, image blur and more intricate atmospheric effects. The issues with supervised methods led researchers to focus on an unsupervised approach [30, 40] where no labels or weights are given and a learning algorithm is left on at its own to find structure in its input. At the moment these methods are not on par with the supervised

category, however, the concept is convincing enough and has potential for future work.

2.2.2 Supervised learning methods

Two types of networks have been proposed. The first type incorporates both feature extraction and matching [23, 41, 42] in one net. The second type performs only one of the two tasks [39, 43]. The most prominent in the first category is FlowNet [23]. Based on the U-Net de-noising auto encoder, FlowNet is the first end-to-end, fully convolutional deep neural network that is trained in a supervised way to produce optical flow using a pair of images. For the first time it is established that optical flow estimation can be posed as a supervised learning problem. Although it is not the state-of-the-art in all aspects, the idea of utilizing convolutional net for optical flow was a breakthrough. It shifted the paradigm of research in this field from traditional to the deep learning approach. The traditionally used datasets (Middlebury, KITTI, MPI-Sintel) were not large enough for training nets. This limitation led the authors to create a synthetic 2D dataset with random background. FlowNet achieved competitive accuracy at frame rate of 5 to 10 frames per second.

2.2.3 Follow-up work

Many modifications were proposed to FlowNet to enhance efficiency and reduce the model size. This includes a rotationally invariant architectures [44], a 3D convolutional network [45] and a light weight convolutional net [26].The most pertinent work following FlowNet is a coarse-to-fine idea of variational methods by Ranjan and Black known as SPyNet [25]. A convolutional network is learned at each level of pyramid to compute optical flow. The computed flow is used to warp the second image to the first image of the next level and so on. SPyNet uses less model parameters with higher accuracy than FlowNetC [23] and lower than FlowNetS [23].

FlowNet2 [24] is an improved version of FlowNet and performs on par with many state of the art, though a little slower than the original FlowNet. By stacking multiple networks, adding warping of the second image with intermediate optical flow and using sub networks for small displacement FlowNet2 decreased the estimation error by more than 50%. Despite its better performance FlowNet2.0 has some limitations. For instance, its model size is much larger (requires over 160 M parameters) than original FlowNet, its multiple modules require sequential training to reduce over-fitting and it takes more computation time than FlowNet. All these factors make FlowNet2.0 unsuitable for mobile and other embedded devices. Sun et al. [41] combined well-established principles of pyramidal processing, warping, and cost volume with deep learning and proposed PWC-Net. It is 17 times smaller and performs better than FlowNet2.0. PWC-Net is the best balance between model size and efficiency.

The above supervised schemes are end-to-end learning. A pair of images is supplied as input and the network computes optical flow by performing both feature extraction and matching. In some methods CNNs have been employed for one of these tasks (not end-to-end learning). PatchBatch [39] used a Siamese net to compute descriptors for each pixel of whole image. The descriptors are fed to the PatchMatch [38] algorithm producing a sparse flow. Finally, the Edge aware optimization of EpicFlow [19] is applied to the sparse flow to obtain a dense motion field. Deep discrete flow [46] also used Siamese net to learn features but optical flow is obtained by discrete optimization.

2.2.4 Unsupervised learning methods

At present, supervised methods are the most successful category of deep learning methods for optical flow with respect to accuracy and efficiency. However, these schemes suffer heavily and fail if a sufficient amount of ground-truth data is not available. Secondly, the network trained under one situation may not work well in other varying conditions. These restrictions of data-driven schemes lead researchers to unsupervised methods. These are knowledge-driven methods, able to train neural nets using unlabeled image pairs to compute optical flow. Although, their performance is not on par with that of the supervised schemes, the approach looks promising and is gaining attention with gradual improvement in performance. The unsupervised network proposed by [47] is based on the classical constraints without regularization. Its loss function is differentiable with respect to the unknown flow field and allows the back-propagation of the error to the previous layers. The loss function proposed by [48] learns optical flow in an unsupervised end-to-end manner. It combines a data term utilizing brightness constancy with a smoothness term and models the expected variation of optical flow. Its overall accuracy remained below that of the original FlowNet [23], except for real images of KITTI (where 100% ground truth is not available). This network [48] was extended by Meister et al. [40] by introducing an unsupervised loss based on occlusion-aware bidirectional optical flow. The model also applies an unsupervised loss to FlowNetC [23] to learn bidirectional flow. The final output is obtained by iterative refinement through multiple networks of FlowNet stacked together. The Model proposed by [49] learns optical flow with proxy ground truth produced by classical methods in an unsupervised manner. The framework suggested by [50] defines the loss as the photometric error between warped

feature map from input image and the target image. The unsupervised neural nets [51] deals with occlusion and large displacement. For the first issue they combined the occlusion map caused by motion with the loss function. For the second problem, the model suggested three additions: a novel warping strategy for large motion learning, supplementary warped inputs during decoder stage, and adopting histogram equalization and channel representation for flow computation.

2.2.5 Advantages and limitations of deep learning methods

Deep learning methods rely on quality and quantity of the labeled training dataset. This is a major issue for deep learning schemes because for real scenes it is extremely cumbersome to obtain labeling. The other problem is the hazardous overfitting due to millions of parameters contained by neural nets. This issue along with large memory trail adversely affects the learning efficiency.

An important aspect is the balance between accuracy and size of deep learning architectures. The design of network is a basic factor behind its efficiency. FlowNet2 [24] stacked multiple modules of FlowNet (Fig. 3). This improved accuracy but also demanded more than 150 million parameters. Large networks tend to consume high memory due to the extended number of parameters used. On practical grounds it is not very functional. SPyNet [25] is comparatively low in both ways. PWC-Net [27] exhibited the best balance between accuracy and size. Many follow up researchers adopted PWC-Net with improved performance [29, 31]. Another important aspect that affects the performance of deep learning methods is the kind of image sequences being used. Traditional and modern researchers have used different datasets. Middlebury [13], Sintel [14], KITTI [15] and Flying Chairs [23] are the most commonly used datasets. Each set has a unique challenge for creating the most accurate system. The Middlebury dataset has only eight frame pairs with ground truth data. The KITTI dataset has around fifty times as many samples as that of Middlebury, with ground truth data, but the images are not labeled densely. MPI-Sintel datasets provided over 1,041 training samples with ground truth data. However, training deep learning models requires several thousands of parameters and a large number of marked samples. All major datasets are still relatively small to be used in a deep learning environment. Further, the small datasets often lead to an increase in the direct training difficulty and over-fitting, which results in reduced accuracy. Due to the reasons above, the Flying Chairs [23] and Freiburg[52] datasets were designed specifically to be used in deep learning schemes.

We compare learning based and traditional methods by their relative advantages and disadvantages. Firstly, deep learning schemes are more efficient in extracting images features to be used for optical flow estimation. This is because of the multi-layer architectures of these methods that allows them to extract more abstract, deeper, and multiscale features. Secondly, these methods avoid the disadvantages of hazardous and complex optimization of the traditional schemes. In credit to the introduction of stochastic minimization of loss function, deep learning methods can better model the intricate, non-linear transformations of the input images [51]. A very important advantage of deep learning schemes is their running speed. Generally, these methods run in real time. On the other hand, traditional methods with similar accuracy take a much longer time (Fig. 4). This makes them impractical for mobile and other embedded devices.

Deep learning schemes also have a number of drawbacks. One of the major aspects affecting their performance is the quality and size of the labelled dataset. This is because, the parameters of convolutional networks are learned from training data. For real image sequences, it is very challenging to obtain dense ground truth labeling. Researchers relied on synthetically rendered,

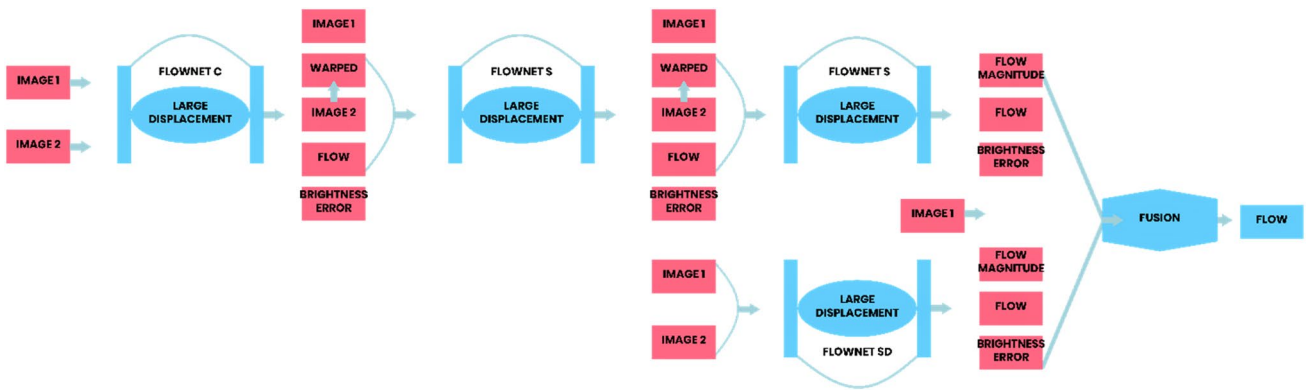
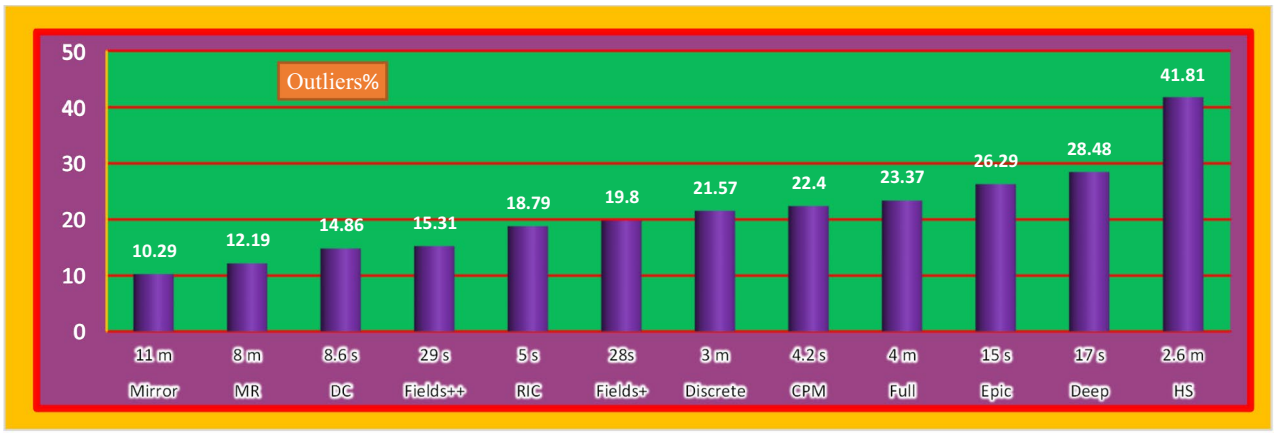
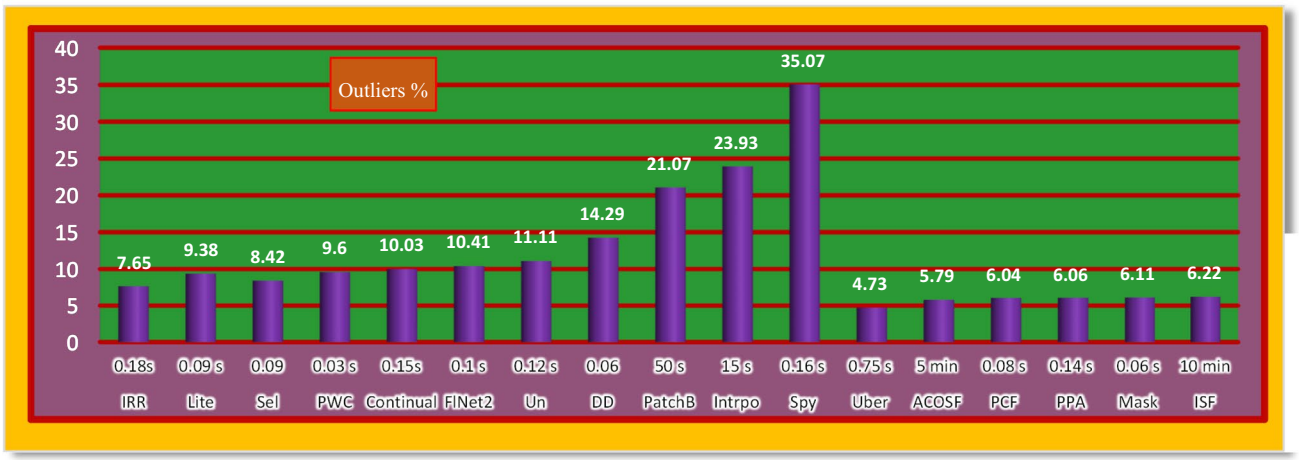


Fig. 3 Multiple modules of Flow Net [23] stacked together in FlowNet2 [24] Architecture



(a) Percentage of outliers and run time for some traditional methods on KITTI 2015 [16] datasets



(b) Percentage of outliers and run time for some deep learning methods on KITTI 2015 [16] datasets

Fig. 4 **a** Percentage of outliers and run time for some traditional methods on KITTI 2015 [16] datasets, **b** Percentage of outliers and run time for some deep learning methods on KITTI 2015 [16] datasets

sequences [23, 52]. Although, the synthetic image sequences exhibit motions, they do not replicate the intricacy of realistic photometric effects such as the blur of different types, noise, brightness variations, atmospheric effects like shadows mist and fog etc. Thus the natural gap between the distribution of synthetic data and real-world scenes is always present and the algorithms trained on synthetic data encounter difficulties when it comes across a more generalized and complex real world image sequences.

Since convolutional networks contain millions of parameters, this aspect along with the difficulty of obtaining suitable training data, provides yet an added disadvantage for deep learning methods: the significant risk of overfitting. This demands that deep learning methods require a large memory footprint. The dependency on the substantial number of parameters, complicates setting an appropriate loss function and may result in a compromised efficiency of the learning process.

3 Evaluation

Optical flow estimation is an extensive field posing ambiguous problems in diverse ways. The major issues making optical flow a complicated subject include occlusions, large displacements, non-rigid motion, discontinuities, mixed pixels, varying illumination, motion and camera blur. In the classic period, Barron [53] established optical flow benchmarks dealing with simple transformations (translation, rotation) and small displacements (Yosemite). Modern researchers created more challenging benchmarks. Below is a discussion of some well-known measures of performance and several datasets adopted by the majority of researchers.

3.1 Measures of performance

An optical flow algorithm finds two dimensional velocity vectors describing the motion field. The degree of success

of a method is measured by reporting its errors. There are mainly two measures for optical flow algorithms:

End point error (EE) This is the fundamental measure and describes the Euclidean distance between.

estimated vectors and the ground truth vectors: $EE = \left|(V - V_g)\right| = \sqrt{(u - u_g)^2 + (v - v_g)^2}$

$V = (u, v)$ is the estimated vector and $V_g = (u_g, v_{g,})$ is the ground truth vector.

Angular error (AE) Primarily used by Barron [53], this is the second most common error measure. Let $(u,v, 1)$ be the extended 3D flow vector and $(u_g, v_g, 1)$ be the ground truth.

AE is the 3D angle: $AE = cos^{-1}[(u.u_g + v.v_g + 1.0)/\sqrt{(u^2 + v^2 + 1.0)(u_g^2 + v_g^2 + 1.0)}]$

Significance AE is appropriate for small displacements and is more inclined to under-estimate large motion. EE is good for large vectors. Advance level sequences [14–16] contain significantly large displacements. This has led most of the modern researchers to report EE instead of AE.

3.2 Datasets for optical flow

Datasets play an important role in computer vision. Image processing domains such as stereo, face and object recognition have challenging datasets. Optical flow was one of the first to introduce standard datasets for quantitative comparisons [53] The improving estimation techniques and modern algorithms demanded an advanced level of datasets for better comparisons of the latest methods. Classic work on optical flow relied on synthetic datasets [53] with Yosmite Sequence being the most well-known. S. Baker presented Middlebury datasets [13] with dense ground truth bringing in new evaluation standards, followed by the path-breaking work of KITTI [15, 16], MPI-Sintel [14] flying chairs [23] and Freiburg [54].

These datasets pose advanced challenges when compared to previously used sequences. The researchers can freely use the dataset for training algorithms and upload the evaluated flow to compare the efficiency of their proposed methods. The online access to these datasets and evaluated optical flow is a prominent feature of the present-day research in the related field. Figure 5 depicts some images from modern datasets used for optical flow estimation. Some of their salient features are discussed below.

3.2.1 Middlebury

It is a leading benchmark to address advanced level problems, covering evaluation at a broader spectrum and wider range of statistical measures [13]. Comprising sub pixel ground truth and ample difficulty, these are realistic synthetic sequences with non-rigid motions, complex scenes and higher texture. However, the motions are small as compared to more advance datasets [14–16]. In contrast to the previously used simple synthetic sequences (Yosemite [53]) these datasets are considerably more challenging which include additional complex scenes, larger ranges, higher realistic texture and independent motion. The sequences are divided into training and testing categories. The ground truth is provided only for the first one. Although the dataset contains advanced level complex motions, most of the motions are small. For training sets, the percentage of the pixels having motion over 20 pixels is less than 3%. The Middlebury are the primary standard datasets posing advanced level of challenges, used by modern algorithms for the estimation of stereo disparity and optical flow. Another important aspect of these datasets is the use of several new measures to test the performance of flow-algorithms. The most important of these measures are average angular error (AE) [53] and endpoint error (EPE).

3.2.2 KITTI

These datasets were created by Geiger in 2012 [15] contain 194 training and 195 test pairs of images with sparse ground truth flow. All images are gray and include complex lighting conditions with large displacements. Later in 2015, Menze annotated the dynamic scenes with 3D CAD models for all vehicles in motion and obtained an extended version, with 200 training and 200 test scenes [16]. The KITTI datasets contain stereo videos of road

Fig. 5 Few frames and the corresponding ground truth from modern datasets

scenes from a calibrated pair of cameras mounted on a car. Ground truth is obtained from real world scenes combining recordings from the camera and a 3D laser scanner. Although, the datasets are real, the density of the ground truth varies from 75 to 90% and underlies the static parts of the scene such as sky or distant objects. For the latest version of KITTI-2015, 3D models of cars were fitted to the point clouds to obtain a denser labeling which also included moving objects. However, the ground truth in these areas is still an approximation. Besides adopting statistical measures described in Sect. 3, KITTI has introduced particular metrics for evaluation to be performed under special circumstances for both versions.

3.2.3 MPI-Sintel

Before introduction of KITTI2015, MPI-Sintel [14] were the largest datasets for optical flow and disparity estimation. Derived in 2012, from an open source 3D-animated film, these datasets are completely synthetic. The original movie frames were modified to pose new challenges for estimation methods. With a resolution of 1024×436, the scenes are designed to be strictly realistic with fog and added motion blur. These datasets provide a naturalistic video sequences containing flow fields, motion boundaries, unmatched regions, and image sequences. The complete data consists of 23 training sequence 1064 frames and 12 testing sequences with 564 frames. Ground truth is available for the training set only. For the modern algorithms, Sintel datasets play a vital role as they provide dense ground truth with adequate occlusion for both small and large displacements. The datasets provide three versions: albedo, clean and final. Albedo is the simplest set and contains no additional effects. The clean version introduces negligible variations to illumination while the final version adds more intricate features such as atmospheric effects, fog, shadows and blur of different types. Researchers commonly use only clean and final versions.

3.2.4 Flying chairs

These images were designed specifically for training Convolutional Neural Networks (CNNs) used in deep learning methods to evaluate optical flow. Created by the application of affine transformations to real images and synthetically rendered chairs, the dataset contains 22,872 image pairs; 22,232 training and 640 test samples according to the standard evaluation split. Dosovitskiy [23] created these datasets to train his convolutional network for optical flow estimation. These datasets are large, do not contain any 3D motions, and hence are limited to single-view optical flow.

3.2.5 Freiburg-Berkeley

This is the latest and the largest data collection for optical flow, stereo and scene flow evaluation [52]. Containing 34,801 stereo training frames and 4248 test frames in 960 X 540 resolutions, these datasets are synthetically produced by 3D suit blended [54] mainly to be employed in deep learning schemes. The first large-scale datasets to enable training and evaluation of scene flow methods consists of three subsets as described below. Table 1 gives salient features of the famous dataset used by modern researchers for optical flow estimation.

4 Developments on major challenges

The majority of optical flow methods are based on brightness constancy and smoothness assumptions. Unfortunately, both are not perfect photometric expressions in practice for many real scenes in motion. A moving light source in a rigid scene will produce brightness variations without moving any object. Similarly, smoothness constraint may not be very accurate on real grounds especially in case of discontinuities where objects occlude one another. Hence all algorithms based on smoothness constraints face difficulties over regions containing fragmented occlusions. The major issues causing erroneous outputs for many algorithms include: outliers, discontinuities, large displacements, occluded edges and insufficient texture.

Table 1 Different datasets and the available ground truth

Datasets	Frame Pairs	Frames with ground truth	Ground truth density (%)
Middlebury	72	8	100
KITTI-2012	195	194	Approx. 50
KITTI 2015	200	200	75–90
MPI-Sintel	1041	1041	100
Flying Chairs	22,872	22,872	100
Freiburg	34,801	34,801	100

In this section we take a look at these challenges and some proposed formulations as how to overcome these issues.

4.1 Outliers

The global methods [33] coupling brightness constancy and spatial smoothness in a variational framework are appropriate for small motions, but face difficulties for extreme values. Outliers impart large errors to estimation. Modern methods have used a robust objective function in which the forward–backward flow is estimated by a bi-directional consistency check and interpolation is applied into the outliers pixels as post processing [20, 21, 32, 36, 39, 46].The same approach with a competent interpolation scheme has been applied by EpicFlow [19] producing convincing outputs. However, basic restrictions at post processing are still expected. EpicFlow [19] used state-of-the art descriptor matching, but possesses a major disadvantage of being sparse. Contrarily, the data based techniques like approximate nearest neighbor field (ANNF) [55] produces dense field but suffers from outliers because they do not incorporate regularization. Among traditional methods, the best algorithm to address this issue was proposed by FlowFields [21]. It replaced the sparse descriptor matching of EpicFlow with a dense corresponding field approach purely based on search strategy without applying any regularization, smoothing or additional data term.

4.2 Discontinuities

Flow discontinuities can impart extensive errors, particularly for the methods based on nearest neighborhood. Since these methods impose spatial or temporal continuity, these assumptions are generally violated at surface boundaries known as motion discontinuities [55]. Classically this issue was addressed by regarding motion discontinuities as outliers and discarding them. Spotting discontinuous regions accurately, with high precision has been the focus of research and several formulations were proposed to the basic HS [33] model. Classical researchers replaced quadratic regulariser by oriented smoothness constraints to prevent blurring across boundaries [53]. Later [56] came up with a heuristic modification to this approach by performing less smoothing close to the image boundaries and determining the smoothing along and across boundaries.

4.3 Large displacements

Large displacement is a fundamental area of concern, responsible for the failure of many optical flow algorithms. It occurs as a result of motion of an object moving at a high speed or due to a low frame-rate. The majority of the varitional algorithms fail to tackle with large displacement because the energy function may be trapped into an incorrect local minimum. The proposed solution can further lead to higher erroneous outcomes due to iteration schemes, which is an essential part of such techniques. To handle large displacements almost all variational methods use coarse-to-fine framework schemes. Among traditional methods, the most prominent work in this regard was done by T. Brox [34] who integrated the variational methods with coarse-to-fine scheme and also provided a theoretical proof of his warping method. Earlier, the coarse-to-fine strategies were used on empirical basis. The method is robust to noise and demonstrates highly accurate results with smaller angular errors. Some methods attempted to solve the large displacement problem without coarse to fine scheme. F. Steinbrucker [35] suggested a quadratic relaxation scheme that does not imply coarse to fine but the algorithm is computationally expensive because it is based on search for candidate correspondences.

Modern researchers paid special attention to the large displacement problem and suggested many novel schemes as probable solutions to this major challenge of optical flow estimation [17, 19, 20, 36, 55, 57]. The most successful solutions are based on feature detection and descriptor matching techniques. LDOF [17] was the first scheme to use local descriptors for dense optical flow. By combining the feature detection and interpolation techniques with optical flow framework LDOF solves the large motion problem substantially more adequately than previous methods. DeepFlow [18] blended matching with variational setup building a multi-stage architecture and interleaving convolutions with max-pooling similar to convolutional neural nets. EpicFlow [19] uses dense matching and a powerful interpolation scheme. NNF [55] used approximate neighbor field and segmentations. CPM [36] merges nearest neighbor field search with coarse-to-fine. S2F-IF [57] combines sparse matching, interpolation and regularizer to perform multi scale matching. Flowfields [21] proposed a dense corresponding field method which outperformed sparse descriptor matching techniques and Discrete Flow [20] applies discrete optimization with sub-pixel refinement.

4.4 Varying illumination

The outdoor machine vision applications, such as traffic monitoring [9] and autonomous driving [5] require robust solutions capable of handling situations under varying light conditions. Optical flow algorithms yield good results for Lambertian surfaces at constant illumination and for the objects moving under homogeneous brightness conditions. However, many weather-linked factors such as

clouds, variable sunshine and fog can affect parts of an image in different ways. Numerous approaches have been taken into consideration in handling illumination changes. Classical local and global methods tackled this issue within the regularization framework requiring many parameters for illumination components to be determined in advance. However, this reduces the applicability of these methods. In recent years, the focal point of research was robustness against large displacement and occlusions; there is no prominent research that specifically focuses on illumination changes. Literature-wise, the most robust methods for illumination changes are Census Transform and the Rank transform [58]. These methods collect signatures from the intensity values in a neighboring patch of the pixel under consideration. However, their key weakness is the reduced output accuracy due to partially lost information in obtaining signatures.

4.5 Lack of texture

The regions with insufficient texture are the main contributors for error estimation in most techniques especially for the methods based on key-point detection and matching. Almost all major methods using feature matching suffer from this common weakness. This issue is well-addressed by the methods based on patch-matching, nearest neighbor field and segmentation [21, 36, 55]. These methods, instead of collecting information from the texture of a particular patch, take into account the motion information from the neighboring patches, making the method robust against texturelessness. The algorithms proposed by [21, 36] performed well for large displacement on MPI-Sintel [14]. The scheme introduced by [55] outperformed for small displacements such as Middlebury [13].

4.6 Occlusions

Occlusions are one of major challenges for modern algorithms that remained unresolved. Occlusions occur when multiple objects in a complex scene move with different displacements and overlap one another in consecutive frames. Violating the data conservation, it leads to errors for multi-frame methods because of the partially lost information. Different approaches dealing with occlusions include: the bidirectional inconsistency check, image warping, data constancy violation and image segmentation. The pixel-wise methods [63] traces the path of every pixel in consecutive images. Using edge detection, matching and warping, [64] introduced an effective interpolation scheme. The segmentation-based methods [55, 65, 66] take uniform motion among small regions to deal with occlusions. Among modern researchers [67] proposed a variational model with a self-adaptive weight energy function and non-local term. In this sense the most powerful schemes are those utilizing occlusions as supplementary evidence to compute optical flow such as MirrorFlow [22], ContinualFlow [28], SelFlow [30]. This approach is contrary to the classical methods applying forward/backward inconsistency check and discarding occlusions as outliers. Table 2 depicts the performance of some famous methods for clean and final versions of MPI-Sintel datasets. The schemes paying special attention to occlusion handling are at the top on both fronts followed by those tackling with large displacement and other issues.

Table 2 Performance of prominent methods in terms of EPE on clean and final versions of MPI-Sintel [14] dataset. Results on the final version are worse than clean because final image sequences present complex and intricate features such as fog, mist, shadows and blur of different types. Among all major challenges faced by modern optical flow algorithms, occlusion handling is one of the top most problems. Methods paying special attention to this issue have outperformed others on both datasets

Traditional methods	Clean pass	Final pass	CNN methods	Clean pass	Final pass
MR-Flow [37]	2.527	5.376	Continual Flow [28]	3.341	4.528
FlowFields+[21]	3.102	5.707	MFF [29]	3.423	4.566
MirrorFlow [22]	3.316	6.071	SelFlow [30]	3.745	4.262
S2F-IF [57]	3.500	5.417	FlowFieldsCNN [21]	3.778	5.363
DCFlow [46]	3.537	5.119	IRR-PWC [31]	3.844	4.579
RicFlow [59]	3.550	5.620	DeepDiscreteFlow [46]	3.863	5.728
CPM-Flow [36]	3.557	5.960	FlowNet2 [24]	3.959	6.016
DiscreteFlow [7]	3.567	5.960	InterpoNet dm [60]	3.973	5.711
FullFlow [32]	3.601	5.895	PWC-Net [27]	4.386	5.042
FlowFields [21]	3.748	5.810	LiteFlowNet [26]	4.539	5.381
EpicFlow [19]	4.115	6.285	FlowNetC+ft+v [23]	6.081	7.883
PHFlow [61]	4.388	7.423	FlowNetS+ft+v [23]	6.158	7.218
DeepFlow [18]	5.377	7.212	DDFlow [62]	6.176	7.401
NNF-Local [55]	5.386	7.249	SPynet+ft [13]	6.640	8.360
LDOF [17]	7.563	9.116	UnFlow [40]	9.379	10.219

5 Discussion

Brief presentations of the development and key points of optical flow modeling and computation have been discussed. The understanding of issues has significantly increased over last 4 decades, and a large variety of estimation models, and optimization schemes have been proposed and combined together for the most accurate results. However, the selection of the best motion estimator is still highly dependent on several factors. Research has given rise to some queries regarding issues related to the existing datasets and estimation techniques for a generic, robust and real-time motion-estimation algorithm. Hence, we look at these issues with an end note for potential future work ensues.

5.1 Choosing an appropriate method

Though each method has its own benefits and limitations, the majority of successful traditional methods choose variational framework and resort to the energy minimization scheme. From the Middlebury [68], MPI-Sintel [69] and KITTI [70] benchmarks, it is obvious that globally regularized models based on joint estimation, segmentation and approximate nearest neighbor field, can produce improved results and perform well for datasets containing small displacements, moderate intensity changes and piecewise smooth displacements of large structures. Alternatively, for the image sequences offering more challenging situations [14, 15] these techniques remain competitive. Hence, implore further improvements to overcome shortcomings and the foremost problems associated with each method especially regarding the computational time.

5.2 Current and potential challenges

Despite great advancements, many methods presented herewith produce erroneous results and remain elusive in challenging situations such as the large displacements, occlusion associated with large motions and motion boundaries, texture-less regions and the large intensity changes in real environments or due to deformations. The methods that specifically handle large displacements, employ feature matching techniques as a part of their main framework. These methods possess their own intrinsic limitations [17–20]. For instance, in most cases the raw patch features obtained by various matching techniques face difficulties when dealing with large variations in appearances, variable radiation, scaling, rotation and repetitive patterns. Despite many improvements, the output of these strategies is mostly affected by the matching noise, a major problem of methods initialized by sparse descriptor matching. It is proposed that these methods need to focus on improving the interpolation schemes. The traditional coarse-to-fine strategy for variational methods for large displacements and the attempts to combine the independent results of feature matching with large displacements usually fail because of the matching errors and noise produced. To overcome these limitations, a potential research area can be the discrete optimization coupled with computationally powerful evaluation schemes. The methods based on data, patch match and nearest neighborhood [21, 55] produce dense motion field, however they have the added disadvantage of being outliers prone. For future work, the methods coupling approximate nearest neighbor field with segmentation need further improvement on regularization, smoothing and filtering techniques to find inliers and avoid extreme values in the data.

5.3 Improving deep learning schemes

The majority of traditional methods outperforming on public benchmarks suffer from heavy computational cost because of the optimization involved in the sparse to dense interpolation. Deep learning methods and the use of GPUs have shown substantial improvements to overcome the run-time problem. Very recently, the methods integrating popular strategies from traditional schemes with deep learning frameworks have outperformed the non-learning category. However, a persistent issue of most of the deep learning schemes is the non-adaptability without retraining across different datasets of varying properties. All supervised deep learning methods require extensive data training whether they compute the optical flow in an end-to-end manner [24] or as deep learning for matching cost computation [46]. There is not a single approach that is versatile and flexible enough to establish well, across different datasets without tuning.

An important aspect in deep learning methods is the size of the network and the number of parameters needed to learn. Large networks consume significant energy and time for learning. Many deep learning methods still have room for further reduction in size. Methods using spatio-temporal filters could possibly achieve this by compressing these filters. This can be accomplished by decreasing filter dimensions or by filter separation. With reduced size, deep networks can be fitted on mobile devices. Exploring more applications and implementation in this field can be a potential future area of research.

Further the results of deep learning methods are not directly comparable with those utilizing hand-crafted features because deep learning methods mostly make use of synthetically prepared datasets [24] which are not a true representative of real life scenarios. Contrarily, the traditional

non-learning methods do not have these limitations. Besides improving the network architectures, the upgrading and expansion of existing datasets to a point close to the real-life challenges can be a prospective area of creativity and a future investigation for deep learning researchers.

The well-known problems discussed above need more future research. It is hoped that with further progress these issues will be better targeted, making the optical flow methods more useful for registration and correspondence problems.

6 Conclusion

Optical flow makes up an important part for many image processing techniques. In presenting its basic principles and highlighting the key features of traditional and deep learning methods a better understanding is accomplished. Two major classes have been compared in terms of their benefits and limitations. Also error measures, and advance level datasets were discussed with their salient features. Moreover, the major current and potential challenges faced by modern algorithms have been discussed along with their proposed solutions. This comprehensive and crisp investigation of the contemporary methods of optical flow can be helpful for present and future research in computer vision.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Li W, Cosker D, Brown M (2016) Drift robust non-rigid optical flow enhancement for long sequences. J Intell Fuzzy Syst 31(5):2583–2595
2. C. Godard, P. Hedman, W. Li, and G. J. Brostow, "Multi-view reconstruction of highly specular surfaces in uncontrolled environments," Proc. - 2015 Int. Conf. 3D Vision, 3DV 2015, pp. 19–27, 2015.
3. M. Khalid, L. Penard, and E. Memin, 2017 Application of optical flow for river velocimetry. Int Geosci Remote Sens Symp. pp. 6243–6246
4. H. Wang, Q. Li, K. Ji, 2015 "The application of optical flow field technology in solar images," 8th International Conference on Intelligent Networks and Intelligent Systems no. x, pp. 86–89
5. Chao H, Gu Y, Napolitano M (2014) A survey of optical flow techniques for robotics navigation applications. J Intell Robot Syst Theory Appl 73(1–4):361–372
6. Hermann S, Werner R (2014) "High accuracy optical flow for 3D medical image registration using the census cost function", Psivt 2013. LNCS 8333:23–35
7. Abdel-Nasser M, Moreno A, Rashwan HA, Puig D (2017) Analyzing the evolution of breast tumors through flow fields and strain tensors. Pattern Recognit Lett 93:162–171
8. Weibel T, Daul C, Wolf D, Rösch R, Guillemin F (2012) Graph based construction of textured large field of view mosaics for bladder cancer diagnosis. Pattern Recognit 45(12):4138–4150
9. Kastrinaki V, Zervakis M, Kalaitzakis K (2003) A survey of video processing techniques for traffic applications. Image Vis Comput 21(4):359–381
10. Ren G, Li W, O'Neill E (2016) Towards the design of effective free-hand gestural interaction for interactive TV. J Intell Fuzzy Syst 31(5):2659–2674
11. Ranftl A, Alonso-Fernandez F, Karlsson S, Bigun J (2015) A real-time adaboost cascade face tracker based on likelihood map and optical flow. IEEE Trans Inf Forensics Secur. 6(6):468–477
12. M. Jain, H. Jegou, and P. Bouthemy, 2013 "Better exploiting motion for better action recognition," Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit, 2555–2562.
13. Baker S, Scharstein D, Lewis JP, Roth S, Black MJ, Szeliski R (2011) A database and evaluation methodology for optical flow. Int J Comput Vis 92(1):1–31
14. Butler DJ, Wulff J, Stanley GB, Black MJ (2012) A naturalistic open source movie for optical flow evaluation. In: ECCV, Part IV, LNCS 7577, 2012, pp 611–625
15. A. Geiger, P. Lenz, and R. Urtasun, 2012 "Are we ready for autonomous driving? the KITTI vision benchmark suite," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 3354–3361
16. M. Menze and A. Geiger, 2015 "Object scene flow for autonomous vehicles," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 07–12-June, pp. 3061–3070.
17. Brox T, Malik J (2011) Large displacement optical flow descriptor matching in variational motion estimation.pdf. IEEE Trans Pattern Anal Mach Intell. 33(3):500–513
18. P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, 2013 "DeepFlow: Large displacement optical flow with deep matching," Proc. IEEE Int. Conf. Comput. Vis., no. Section 2, pp. 1385–1392.
19. J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, 2015 "EpicFlow: Edge-preserving interpolation of correspondences for optical flow," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 07–12-June, pp. 1164–1172.
20. M. Menze, C. Heipke, and A. Geiger, 2015 "Discrete Optimization for Optical Flow," 37th Ger. Conf. GCPR 2015, vol. i, pp. 16–28.
21. C. Bailer, B. Taetz, and D. Stricker, 2015 "Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation," Proc. IEEE Int. Conf. Comput. Vis., vol. 2015 Inter, pp. 4015–4023.
22. J. Hur and S. Roth, "MirrorFlow: Exploiting Symmetries in Joint Optical Flow and Occlusion Estimation," Proc. IEEE Int. Conf. Comput. Vis., vol. 2017-Octob, pp. 312–321, 2017.
23. A. Dosovitskiy et al., 2015 "FlowNet: Learning optical flow with convolutional networks," Proc. IEEE Int. Conf. Comput. Vis., vol. 2015 Inter, pp. 2758–2766.
24. E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, 2016 "FlowNet 2.0: evolution of optical flow estimation with deep networks,"

25. Ranjan A, Black MJ (2016) Optical flow estimation using a spatial pyramid network. Cvpr 2017:4161–4170
26. Hui T-W, Tang X, Loy CC (2018) LiteFlowNet: A Lightweight Convolutional Neural Network for Optical Flow Estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, pp 8981–8989
27. Sun D, Yang X, Liu M-Y, Kautz J (2017) PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp 8934–8943
28. Neoral M, Šochman J, Matas J (2018) Continual Occlusions and Optical Flow Estimation. In book: Computer Vision-ACCV, pp 159–174. https://doi.org/10.1007/978-3-030-20870-7_10
29. Z. Ren, O. Gallo, D. Sun, M. H. Yang, E. B. Sudderth, and J. Kautz, 2019 "A fusion approach for multi-frame optical flow estimation," Proc. - 2019 IEEE Winter Conf. Appl. Comput. Vision, WACV 2019, pp. 2077–2086.
30. Liu P, Lyu M, King I, Xu J (2019) SelFlow: Self-Supervised Learning of Optical Flow. In: IEEE CVPR, pp 4571–4580
31. J. Hur and S. Roth, 2019 "Iterative Residual Refinement for Joint Optical Flow and Occlusion Estimation," [cs.CV] 10.
32. Q. Chen and V. Koltun, 2016 "Full Flow: Optical Flow Estimation By Global Optimization over Regular Grids," Proc. IEEE Conf. Comput. Vis. Pattern Recognit.
33. Horn BKP, Schunck BG (1981) Determining optical flow. Artif Intell 17(1–3):185–203
34. T. Brox, N. Papenberg, and J. Weickert, 2004 "High Accuracy Optical Flow Estimation Based on a Theory for Warping," Comput. Vis. - ECCV 2004, vol. 4, no. May, pp. 25–36.
35. F. Steinbrucker, T. Pock, and D. Cremers, 2009 "Large Displacement Optical Flow Computation without Warping,"The 12th International Conference on Computer Vision, p.1609–1614.
36. Y. Hu, R. Song, and Y. Li, 2016 "Efficient coarse-to-fine patchmatch for large displacement optical flow," Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 5704–5712.
37. J. Wulff, L. Sevilla-Lara, and M. J. Black, 2017 "Optical flow in mostly rigid scenes," Proc.30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 6911–6920.
38. C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein, 2010 "The generalized PatchMatch correspondence algorithm," Lect. Notes Comput. Sci. vol. 6313 LNCS, no. PART 3, pp. 29–43.
39. D. Gadot and L. Wolf, 2016 "PatchBatch: A Batch Augmented Loss for Optical Flow," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Dec. pp. 4236–4245.
40. S. Meister, J. Hur, and S. Roth, 2017 "UnFlow: Unsupervised Learning of Optical Flow with a Bidirectional Census Loss," [cs.CV] 21.
41. D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, 2018 "PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume," [cs.CV] 25 Jun
42. A. Ranjan and M. J. Black, 2017 "Optical Flow Estimation using a Spatial Pyramid Network,"CVPR 2017, pp. 4161–4170
43. T. Schuster, L. Wolf, and D. Gadot, 2017 "Optical flow requires multiple strategies (but only one network)," Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 6921–6930.
44. D. Teney and M. Hebert, 2017 "Learning to extract motion from videos in convolutional neural networks," Lect. Notes Comput. Sci. vol. 10115 LNCS, pp. 412–428.
45. D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, 2016 "Deep End2End Voxel2Voxel Prediction," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., pp. 402–409.
46. F. Güney and A. Geiger, 2017 "Deep discrete flow," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10114 LNCS, pp. 207–224.
47. A. Ahmadi and I. Patras, 2016 "Unsupervised convolutional neural networks for motion estimation," IEEE International Conference on Image Processing (ICIP)
48. J. J. Yu, A. W. Harley, and K. G. Derpanis, 2016 "Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness," Lect. Notes Comput. Sci. vol. 9915 LNCS, pp. 3–10.
49. Y. Zhu, Z. Lan, S. Newsam, and A. G. Hauptmann, 2017 "Guided Optical Flow Learning," CVPRW, February.
50. Z. Ren, J. Yan, B. Ni, B. Liu, X. Yang, and H. Zha, 2017 "Unsupervised Deep Learning for Optical Flow Estimation," Proc. 31th Conf. Artif. Intell. (AAAI 2017), no. Hollingworth 2004, pp. 1495–1501.
51. Y. Wang, Y. Yang, Z. Yang, L. Zhao, P. Wang, and W. Xu, 2018"Occlusion Aware Unsupervised Learning of Optical Flow," Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit, pp. 4884–4893.
52. N. Mayer et al., 2016 "A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation," IEEE Conference on Computer Vision and Pattern Recognition pp. 4040–4048.
53. Barron JL, Fleet DJ, Beauchemin SS (1994) Performance of optical flow techniques. Int J Comput Vis 12(1):43–77
54. Blender. https://www.blender.org
55. Z. Chen, H. Jin, Z. Lin, S. Cohen, Y. Wu, 2013 "Large displacement optical flow from nearest neighbor fields," Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit, pp. 2443–2450.
56. Alvarez L, Deriche R, Papadopoulo T, Sánchez J (2007) Symmetrical dense optical flow estimation with occlusions detection. Int J Comput Vis 75(3):371–385
57. Y. Yang, S. Soatto, 2017 "S2F: Slow-to-fast interpolator flow," Proc. - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017, pp. 3767–3776.
58. D. Hafner, O. Demetz, J. Weickert, 2013 "Why is the census transform good for robust optic flow computation?," Lect Notes Comput Sci, vol. 7893 LNCS, pp. 210–221.
59. Y. Hu, Y. Li, R. Song, 2017 "Robust interpolation of correspondences for large displacement optical flow," Proc. - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 4791–4799.
60. S. Zweig and L. Wolf, "InterpoNet, a brain inspired neural network for optical flow dense interpolation," Proc—30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017, pp. 6363–6372.
61. J. Yang and H. Li, 2015 "Dense, accurate optical flow estimation with piecewise parametric model," Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit, vol. 07–12-June, pp. 1019–1027.
62. P. Liu, I. King, M. R. Lyu, and J. Xu, "DDFlow: learning optical flow with unlabeled data distillation," The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19) pp: 8770–8777.
63. Mahajan D, Huang F-C, Matusik W, Ramamoorthi R, Belhumeur P (2009) Moving gradients. ACM Trans Graph 28(3):1
64. Stich T, Linz C, Wallraven C, Cunningham D, Magnor M (2011) Perception-motivated interpolation of image sequences. ACM Trans Appl Percept 8(2):1–25
65. Unger M, Bischof H (2012) Joint motion estimation and segmentation of complex scenes with label costs and occlusion modeling. IEEE CVPR 2:1878–1885
66. Chen K, Lorenz DA (2012) Image sequence interpolation based on optical flow, segmentation, and optimal control. IEEE Trans Image Process 21(3):1020–1030
67. Zhang C, Chen Z, Wang M, Li M, Jiang S (2017) Robust non-local TV-L1 optical flow estimation with occlusion detection. IEEE Trans Image Process 26(8):4055–4067
68. http://vision.middlebury.edu/flow/eval/
69. http://sintel.is.tue.mpg.de/results
70. http://www.cvlibs.net/datasets/kitti