IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 6, 2019, accepted January 3, 2020, date of publication January 20, 2020, date of current version January 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2968180

Improved Discrete Optical Flow Estimation With Triple Image Matching Cost

FEI YANG[1,2], YONGMEI CHENG[1], JOOST VAN DE WEIJER[2], AND MIKHAIL G. MOZEROV[2], (Member, IEEE)

[1]Key Laboratory of Information Fusion Technology, Northwestern Polytechnical University, Xi'an 710072, China
[2]Computer Vision Center, Department of Informatics, Universitat Autónoma de Barcelona, 08193 Barcelona, Spain

Corresponding author: Fei Yang (fyang@cvc.uab.es)

This work was supported in part by the Spanish Project (MINECO/FEDER) under Grant TIN2015-65464-R and Grant TIN2016-79717-R, in part by the COST Action IC1307 iV&L Net (European Network on Integrating Vision and Language), and in part by the European Cooperation in Science and Technology (COST). The work of Fei Yang was supported by the Chinese Scholarship Council (CSC) under Grant 201706290127.

ABSTRACT Approaches that use more than two consecutive video frames in the optical flow estimation have a long research history. However, almost all such methods utilize extra information for a pre-processing flow prediction or for a post-processing flow correction and filtering. In contrast, this paper differs from previously developed techniques. We propose a new algorithm for the likelihood function calculation (alternatively the matching cost volume) that is used in the maximum a posteriori estimation. We exploit the fact that in general, optical flow is locally constant in the sense of time and the likelihood function depends on both the previous and the future frame. Implementation of our idea increases the robustness of optical flow estimation. As a result, our method outperforms 9% over the DCFlow technique, which we use as prototype for our CNN based computation architecture, on the most challenging MPI-Sintel dataset for the non-occluded mask metric. Furthermore, our approach considerably increases the accuracy of the flow estimation for the matching cost processing, consequently outperforming the original DCFlow algorithm results up to 50% in occluded regions and up to 9% in non-occluded regions on the MPI-Sintel dataset. The experimental section shows that the proposed method achieves state-of-the-arts results especially on the MPI-Sintel dataset.

INDEX TERMS Motion estimation, optical flow, matching cost, multi-frames optical flow.

I. INTRODUCTION

Optical flow estimation is important for a large variety of computer vision applications, such as 3D scene reconstruction, autonomous driving systems and robotics. Thus, optical flow can be considered as one of the fundamental problems of computer vision. Originally, the optical flow methods are based on the assumptions of brightness constancy and spatial smoothness [1], [2]. Although there is a long research history, accurate and robust optical flow is still an open problem due to illumination changes, large displacement, blur, texture-less regions and occlusions.

Recently, several new approaches [3]–[6] leverage end-to-end convolutional neural networks [7] to take an important step forward in optical flow estimation, and results are close to state-of-the-art. However, these networks without a special architecture for optical flow estimation can realize their full potential only with adequate training data and appropriate training arrangements [8]. For example, PWC-Net [6] gets state-of-the-art results with a new neural network architecture by embedding several classical principles: pyramid, warping and cost volume.

In this paper, we focus our efforts to improve the part of neural network architectures that is based on classic methods. Several typical methods use an initialization by approximate nearest neighbor fields (ANNF) [9]–[12] or sparse descriptor matching [13], they leverage edge-preserving interpolation techniques [14], [15] to get final dense optical flow. High quality correspondence is the key for dense optical flow estimation.

Early optical flow methods make two assumptions: the optical flow motion vector field belongs to the set of

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

continuous values; and the magnitude of the motion vector values is relatively small. Thus, the problem of large motion vectors arises. Note that the same problem exists in stereo matching and the most successful algorithms rely on discrete inference, where all possible discrete disparity vectors form the matching search domain. Stereo matching methods achieve an impressive accuracy, and this is why many optical flow estimation algorithms try to exploit the same technique to solve the optical flow problem in the framework of the discrete matching paradigm [16]–[19]. Generally, a stereo matching method pipeline consists of the following steps: matching cost computation, cost aggregation or optimization and post-processing refinement. Unfortunately, the straightforward application of this stereo matching scheme to optical flow is very difficult due to the huge size of the discrete 2D motion vector domain in comparison to the 1D stereo problem [19]. Recent progress in parallel calculation architectures shows that processing on the non-restricted cost volume is feasible and that the regular structure of this volume allows the use of global optimization techniques [20].

The main part of the above pipeline is the cost volume formation. A matching cost or a dissimilarity measure is an essential part of the correspondence problem that, in turn, is a fundamental problem in computer vision. Thus, calculation of the cost volume in stereo matching and discrete optical flow is a very important sub-problem [21]. There are two kinds of cost calculation approaches: a per-pixel and an area based dissimilarity measure estimation. The per-pixel dissimilarity measure usually is the Euclidean distance in the RGB color space between two image matched values or the same distance between the gradients. The robust per-pixel measure is reported [17], [18] when distances between gradients and values are combined in one measure. Early methods that exploited area based cost models calculated the cost by using a non-parametric transform with a support region such as rank and census [22] or normalized cross correlation [23]. Using a combination of these two costs can significantly improve the result of stereo matching [24]. A patch match approach is proposed in [25], where they used the sum of squared distances to compute an initial matching cost. Consequently, Kong and Tao [25] propose a new cost learning technique, which is theoretically extended by Brown *et al.* [26]. The latest progress in the field of CNN provides a more robust matching cost for stereo [27]–[29] and optical flow [10], [30]. Consequently, the traditional cost computation has been replaced by the CNN based cost in most recent works, and we also include the CNN based framework as a part of our cost calculation process.

Despite progress in the robust matching cost formation there is still one fundamental problem in the matching dissimilarity estimation: the cost uncertainty in the occluded region, due to the lack of a real correspondence between matched pixels in this case. For the standard stereo matching that uses only two images the mentioned problem cannot be solved in a straightforward manner. Fortunately, optical flow methods usually deal with more than two images and in the presented work we show how to handle occlusion problem using three consecutive images in the considered video sequences. Note that the occlusion handling in the cost volume domain improves the solution robustness also in non-occluded regions, because the energy minimization approach is very sensitive to the cost outliers.

Formally, all methods that use more than two images can be considered as related work, however the proposed triple patch match model is fundamentally different from these approaches. Usually, the related work introduces a temporal regularization [31]–[34], a trajectory regularization [35]–[37] or predicts optical flow between previous frames to guide the estimation of the current flow field [38], [39]. Another related work is [40], which proposed a variational model for joint optical flow and occlusion estimation with three frames. Recently, there are several papers that embed a multi-frame optical flow estimation into the convolutional neural network architecture. Maurer and Bruhn [39] proposed an unsupervised online learning approach that estimates a current motion model with multi-frame and provides predicted motion information for forward flow estimation. Janai *et al.* [41] proposed an unsupervised learning method for multi-frame optical flow. They construct past cost volume and future cost volume with three frames and leverage convolutional neural network to reason occlusion. Neoral *et al.* [42] also estimate occlusion masks by introducing the previous frame flow and named it ContinualFlow. Ren *et al.* [43] use a neural network to fuse optical flows of different moments depending on longer-term temporal cues.

In contrast, we propose a new matching cost formation based on two assumptions: most occlusion regions that are invisible in the forward frame image (relative to the current frame) are visible in the backward frame; the forward flow is approximately equal to the negative value of the backward flow. The assumptions allow us to form the composite matching cost as a combination of two independent forward and backward matching costs. We consider the proposed composite cost as the main contribution of the paper. Implementation of our method increases the robustness of the optical flow estimation.

To demonstrate the advantage of the proposed cost formation we incorporate our cost in the pipeline of the state-of-the-art method DCFlow [30] and perform several experiments during each step of the prototype estimation scheme. Consequently we show that the results of the prototype method is improved for results of intermediate steps and for the final estimation of the full pipeline. Our approach considerably increases the optical flow estimation on the MPI-Sintel dataset [44] after the matching cost processing that is the most important part of the proposed pipeline. As a result, accuracy of our estimation without the back flow consistency check increases up to 50% in occluded regions and up to 9% in non-occluded regions relative to the DCFlow algorithm original results. After post-processing steps our estimation results

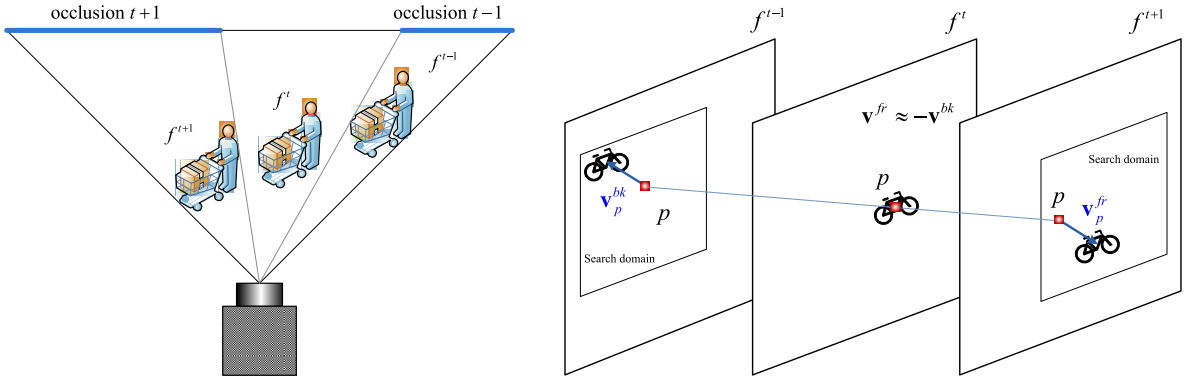


FIGURE 1. Illustration of two main principles for the triple image matching: (a) - supplementing visibility of occluded regions in a triple frame set of a video sequence; (b) - local time optical flow constancy.

achieve state-of-the-art results especially on the MPI-Sintel dataset.

The rest of this paper is organised as follows. In section II we introduce our problem definition. In section III we describe the proposed cost volume formation with triple frame. Our improved optical flow estimation pipeline is described in section IV. In section V we present our experiments. In section VI we conclude our work and plan our further research.

II. PROBLEM DEFINITION

Discrete optical flow estimation belongs to the general matching problem, and in the framework of the global approach the matching problem is formulated in terms of energy minimization with the energy function in the following form:

$$E(v) = \sum_{p \in \mathcal{V}} C_p(v_p) + \sum_{(p,q) \in \mathcal{E}_p} B_{p,q}(v_p, v_q) \tag{1}$$

where set $p \in \mathcal{V}$ corresponds to pixels and set $(p, q) \in \mathcal{E}_p$ to edges of a pixel p neighborhood of an image graph $\mathcal{G} = (\mathcal{E}, \mathcal{V})$; v_p denotes the label of pixel p which belongs to some discrete set of 2D motion vectors $v \in V$ that represents the so called correspondence search region; $C_p(\cdot)$ defines a unary potential which corresponds to the conventional penalty or dissimilarity cost; $B_{p,q}(\cdot, \cdot)$ is a binary potential which defines edge interaction between pixels (p, q). Here we assume that the search region is discrete and rectangular $V = [-v_{max}, -v_{max+1}, \ldots, v_{max-1}, v_{max}]$.

Consequently, the integer solution of the optical flow estimation problem $\mathbf{v}$ should minimize the energy functional in Eq. (1):

$$v_p = \arg\min_{v_p} E(v_p) \tag{2}$$

The binary potential $B_{p,q}$ in Eq. (1) defines the local smoothness of the estimated optical flow v and in our algorithm has the following form:

$$B_{p,q} = \mu \min\left(\left|v_p - v_q\right|, \Delta\right) \tag{3}$$

where μ and Δ the algorithm intrinsic parameters.

As we note in the introduction the choice of the unary potential (cost) in matching tasks is very important. The main contribution of our paper is a new dissimilarity cost formation based on triple image matching. We explain the main concept and motivation of the cost calculation in the next section. However, firstly it is necessary to describe the prototype cost calculation based on a simple convolutional neural network (CNN).

To calculate the cost volume in our pipeline we use the feature extraction CNN network trained by [30]. This small network contains 4 convolutional layers and each layer uses 64 filters. The first three layers are followed by a ReLU layer and output are normalized to produce a unit-length feature. The receptive field of this network or the size of a matched image patch is 9×9, which has proven to be effective for stereo and optical flow estimation. Or formally, the considered CNN transforms the 81D vectors $i_{\mathcal{N}_p}$ that consist of image values in the patch neighborhood relevant to a pixel p into the 64D u_p vectors of the CNN feature space: $u_p = \mathcal{T}_{CNN}(i_{N_p})$.

In turn, the cost is the vector dot-product of two matched pixel features:

$$C(p, v) = 1 - u_p^t u_{p+v}^{t+1} \tag{4}$$

For more details about the used CNN readers can refer the paper [30].

Energy minimization methods have lately attracted much attention in computer vision, especially in the context of image segmentation and optical flow estimation. The first implementations of the energy minimization methods such as belief propagation [45] and graph cuts [46] in stereo matching have provided a significant progress in disparity map estimation. However in our case, where a huge cost volume has to be handled in Eq. (1) the above approaches are computationally demanding. As the trade-off between computational complexity and accuracy of energy minimization we use semi-global matching technique (SGM) [47] to process the cost volume the same as in the DCFlow method [30].

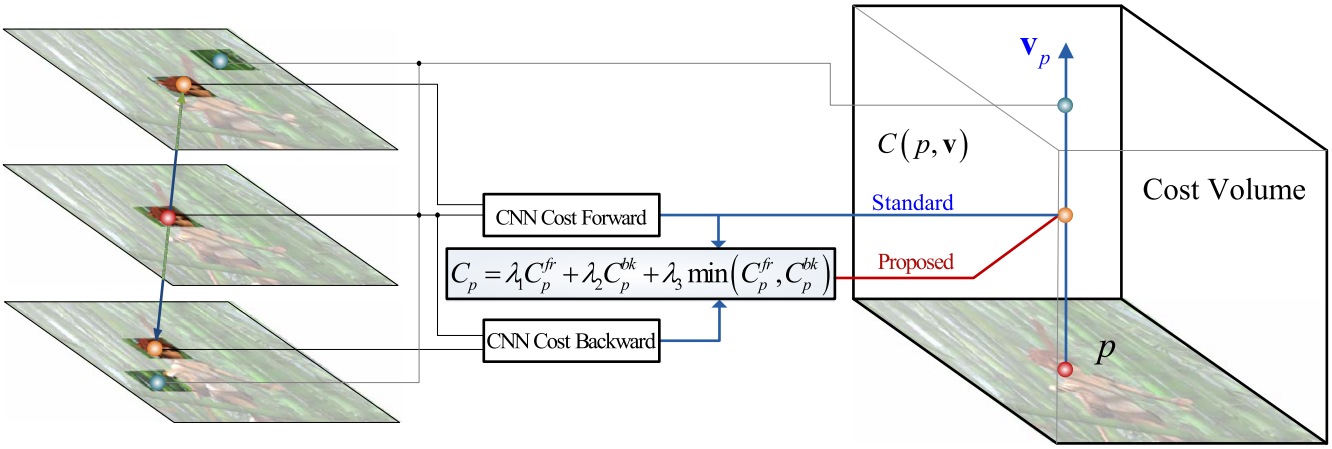


FIGURE 2. Cost volume calculation scheme for the standard CNN based approach and for the proposed triple image matching technique. Both individual costs are unified in one triple image matching cost. Here, the middle image is used twice: for the forward and backward costs calculation.

III. COST VOLUME FORMATION BASED ON TRIPLE IMAGE MATCHING

Our idea to use three constitutive frames for cost calculation is based on two principles: supplementing visibility of occluded regions in a triple frame set of a video sequence and local time optical flow constancy.

The first principle is illustrated in Fig. 1(a). One can see that the occlusion region $t+1$ in the frame f^{t+1} has no corresponded pixels relative to the current frame f^t, but this occluded region is visible in the frame f^{t-1}. The same supplementing visibility exists for the occluded region $t-1$.The illustrated assumption is not a physical law or a strict general observation, however in real world scenarios, those pixels which are visible in a current frame and turn invisible in the next frame are usually visible in the previous frame. Thus, in the set that consists of three consecutive frames there are less pixels in the current frame that have no correspondent pixels in the next or in the previous frames.

The second principle is illustrated in Fig. 1(b). We suppose that the motion vector v_p^{fr}, which corresponds to the forward optical flow direction (to the future) is equal to the negative motion vector $-v_p^{bk}$, which corresponds to the backward optical flow direction (to the past). This principal is a direct consequence of the optical flow framework, and we reformulate it in the cost form rewriting Eq. 4:

$$C^{fr}(p, v) = 1 - u_p^t u_{p+v}^{t+1} = 1 - u_p^t u_{p-v}^{t-1} = C^{bk}(p, -v) \quad (5)$$

It is obvious that the above equality Eq. 5 holds only for non-occluded pixels in the next and the previous frames simultaneously. And for these pixels it is reasonable to make the final cost as a linear combination of the forward and backward costs to make the composite cost more robust:

$$C = \lambda_1 C^{fr} + \lambda_2 C^{bk} \quad (6)$$

However, if Eq. 5 does not hold as illustrated in Fig. 1(a), we assume that one of the costs C^{fr} or C^{bk} is the true cost. Consequently, to avoid ambiguities caused by occlusion, we have to choose the true one. Recall that the energy minimization approach is derived from the maximum a posteriori probability rule with the assumption that the cost of the estimated motion vector is inversely proportional to its probability: $C \propto -\log P$. It means that a lower cost value corresponds to a higher probability. Because we think that this is a good reason to choose the most probable cost as a true cost, consequently, we formalize our paradigm in the presence of occlusion as a minimum choice between correspondent cost values:

$$C = \min\left(C^{fr}, C^{bk}\right) \quad (7)$$

To unify both sets of pixels: occluded and non-occluded, the final cost can be written in the following form:

$$C = \lambda_1 C^{fr} + \lambda_2 C^{bk} + \lambda_3 \min\left(C^{fr}, C^{bk}\right) \quad (8)$$

where linear weights λ_1,λ_2 and λ_3 are our algorithm intrinsic parameters to be optimized and we explain their choice in the experimental section. In Fig. 2 the computational scheme of the composite cost volume is summarized. Also one can see the difference between the proposed cost formation and the standard one.

IV. IMPROVED OPTICAL FLOW ESTIMATION PIPELINE

To demonstrate advantages of our proposed cost formation we design our optical flow estimation pipeline based on triple image matching cost formation (**TIMCflow**), which mainly follows the DCFlow algorithm [30]. In Fig. 3 we depict our main algorithm (red arrow) in parallel with the two-frame DCFlow prototype (black arrow). We demonstrate the result difference between the compared algorithms in all control points (steps) by including the relevant table in the same figure. The compared intermediate results are based on the Sintel training dataset under the EPE of **all** | **noc** | **occ** metrics. Note that numbers in Fig. 3 corresponding to the outlier handling step are not meaningful, because the sparseness density of the algorithms is different.

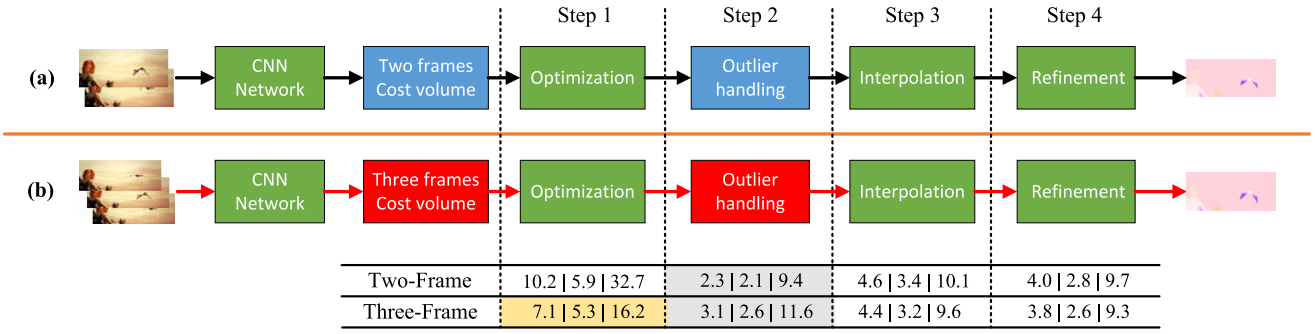


FIGURE 3. The pipeline of two-frame baseline method (a) and our TIMCflow algorithm (b) with the results comparison on MPI Sintel data set after each step.

TABLE 1. Results on the final pass of the MPI-Sintel benchmark for different regions, velocities (s) and distances from motion boundaries (d).

Method	all	matched	unmatched	d0-10	d10-60	d60-140	s0-10	s10-40	s40+
FlowFieldsCNN [10]	5.363	2.303	30.313	4.718	2.020	1.399	1.032	3.065	32.422
CPM-Flow [12]	5.960	2.990	30.177	5.038	2.419	2.143	1.155	3.755	35.592
FullFlow [20]	5.895	2.838	30.793	4.905	2.506	1.913	1.136	3.373	35.136
FlowFields [9]	5.810	2.621	31.799	4.851	2.232	1.682	1.157	3.739	33.890
DCFlow [30]	5.119	2.283	28.228	4.665	2.108	1.440	1.052	3.434	**29.351**
EpicFlow [14]	6.285	3.060	32.564	5.205	2.611	2.216	1.135	3.727	38.021
InterpoNet_ff [48]	5.535	2.372	31.296	4.720	2.018	1.532	1.064	3.496	32.633
ours	**5.049**	**2.094**	**29.134**	**4.738**	**1.812**	**1.221**	**0.922**	**3.226**	29.926
PWC-Net [6]	5.042	2.445	26.221	4.636	2.087	1.475	0.799	2.986	31.070
ProFlow [39]	5.017	2.596	24.736	5.016	2.146	1.601	0.910	2.809	30.715
Back2FutureFlow [41]	8.814	5.031	39.647	7.153	4.880	3.904	1.752	5.961	50.725
MFF [43]	4.566	2.216	23.732	4.664	2.017	1.222	0.893	2.902	26.810

The cost volume formation of the scheme Fig. 3 is described in Section III, and in the experimental section we compare the simple flow results obtained by two different cost formation.

The optimization block of the scheme in Fig. 3 is described in Section II. From Fig. 3 one can see that after the optimization step our three-frame approach considerably outperforms the two-frame prototypes (results highlighted by yellow).

The next step is the occlusion detection and outlier removal. To perform this task most work uses the forward-backward consistency strategy [12], [19], [30]. We also follow this idea, but the problem is that the consistency check procedure usually removes estimated flow values in occluded regions, thus we cannot capitalize advantages of the flow estimation that are achieved on the previous algorithmic step. Consequently we try several different strategies for outlier removal and choose the best that is described in the experimental section.

The next step of our pipeline is the sparse data interpolation, because the output result of the previous steps is the sparse set of estimated values and this set should be interpolated to the dense optical flow. For this purpose we choose the state-of-the-art interpolation method InterpoNet [48]. The motivation behind this choice is that the method can produce good dense optical flow for all kinds of sparse optical flow input, for example FlowField [9], DeepMatch [11], DF [19], CPM [12].

Both two-frame and three-frame pipelines include the same coarse-to-fine procedure based on a continues optimization framework [49] that is the final step of our pipeline.

The performance of the proposed algorithm on MPI Sintel benchmark is confirmed in Table 1, where we compare the proposed algorithm TIMCflow with five discrete optical flow methods: FlowFieldsCNN [10], CPM-Flow [12], FullFlow [20], FlowFields [9] and DCFlow [30], two interpolation methods with discrete optical flow initialization: EpicFlow [14] and InterpoNet [48]. One can see that our method is better than the prototype DCFlow method and also outperforms all compared algorithms.

In addition, we also show the results of several recent learning based optical flow estimation methods in this table: PWC-Net [50], ProFlow [39], Back2FutureFlow [41], MFF [43]

V. EXPERIMENTAL RESULTS

The experiments have been designed to demonstrate the main advantages of the proposed cost formation approach. They are divided into several key parts related to the main algorithm steps in Fig. 3 where:

- the robustness of the proposed triple image matching cost in comparison with standard two-image matching cost is evaluated.
- the advantage of using the proposed cost in the energy minimization part of the pipeline is analyzed

TABLE 2. Cost volume processing results with WTA output: quantitative comparison of two-frames and three-frames using the endpoint error metric for every MPI-Sintel training set sequence. The top part of this table is the all pixels mask, the middle is the non-occluded pixels mask and the bottom is the occluded pixels mask.

Method	average	alley-1	alley-2	ambush-2	ambush-4	ambush-5	ambush-6	ambush-7	bamboo-1	bamboo-2	bandage-1	bandage-2	cave-2	cave-4	market-2	market-5	market-6	mountain-1	shaman-2	shaman-3	sleeping-1	sleeping-2	temple-2	temple-3
3Fs	18.2	2.82	3.82	72.2	58.8	30.2	54.6	5.14	4.70	8.60	5.08	2.74	30.1	19.3	4.28	41.7	21.6	4.13	2.83	2.69	1.99	1.24	11.8	27.6
2Fs	22.6	4.36	5.79	81.4	66.2	36.6	63.2	8.24	6.92	12.7	7.90	4.06	39.3	27.5	6.96	50.3	28.8	6.04	4.56	3.88	2.09	1.30	16.4	35.8
3Fs	15.1	2.38	3.18	66.8	52.4	23.6	48.8	3.97	4.02	6.79	3.92	2.47	22.6	14.6	3.15	33.2	15.0	3.70	2.58	2.55	1.99	1.24	9.61	19.3
2Fs	17.9	3.50	4.73	72.8	56.1	27.2	54.2	6.40	5.74	9.72	5.99	3.61	27.0	19.9	4.84	38.2	20.1	5.32	4.05	3.66	2.09	1.29	12.1	23.6
3Fs	39.5	15.5	40.4	88.0	87.9	62.6	76.5	28.3	19.2	35.0	27.2	8.97	63.7	55.0	26.9	82.8	53.0	18.6	5.94	11.0	1.60	1.51	37.1	61.1
2Fs	58.0	29.8	67.1	109	111	84.2	99.4	50.2	31.9	60.0	44.7	14.9	96.5	86.0	49.04	110	73.3	30.9	10.6	17.4	1.70	1.75	63.2	89.1

TABLE 3. Quantitative comparison of two-frames and three-frames using the endpoint error metric with different masks.

Method	MPI-Sintel			KITTI			Middlebury
	all	noc	occ	all	noc	occ	epe
Two-frame	10.25	5.90	32.74	17.04	7.92	53.57	0.6713
Three-frame	7.18	5.39	16.29	14.87	7.16	43.27	0.6609
Three-frame+	5.99	4.26	14.27	12.14	5.56	36.28	0.5909
Three-frame++	4.65	3.01	12.63	6.00	2.13	21.37	0.2279

(corresponding to Step 1 in Fig. 3) and the results of our additional simplified pipeline of Fig. 5 are reported.

- we motivate our choice for the final outlier handling strategy by analizing the intermediate results after Step 2.
- we compare the results of two different pipelines after flow field interpolation (corresponding to Step 3) and refinement (corresponding to Step 4).
- we report and compare running time of our pipeline in comparison with the two-frame version of our algorithm.

In our experiments we mainly use the final pass of the MPI-Sintel dataset [44] that is a challenging flow evaluation benchmark, which contains long image sequences with large displacements, motion blur, defocus blur and specular reflections. For several additional experiments we also use the KITTI flow 2015 dataset [51] and a part of the Middlebury training dataset [52]. We discuss the results of the proposed algorithm in comparison with state-of-the-art methods on the Sintel dataset. Note that the KITTI dataset differs from the Sintel dataset: the first data set include shading and over-exposure, the second motion blur and dramatic occlusion. As a result, different strategies are necessary to reach state-of-the-art.

A. WTA OUTPUT RESULTS COMPARISON

We perform experiments with the winner takes all (WTA) output for two different cost formation approaches in Fig. 3. In this part, we found that the best results of our approach can be achieved by using $\lambda_1 = 0, \lambda_2 = 0$ and $\lambda_3 = 1$ as parameter setting in Eq. 8. In Table 2 the comparison between two different cost calculations is shown by using the final pass of the Sintel training data. One can see that our triple image matching cost produces more accurate results and improves the accuracy of the standard cost calculation technique in the occluded area by 32% and by 16% in the non-occluded region. The comparison results confirm the ability of the proposed cost to handle occlusion. Note that the proposed cost formation is more robust than the standard two-image matching cost and that is confirmed by the results in the non-occluded region.

B. DISCRETE FLOW RESULTS COMPARISON

The next experiments are performed to show the advantage of using the proposed cost in the energy minimization part of the pipeline (Step 1 in Fig. 3). Here we use the SGM approach to minimize energy of the cost volume for three optical flow datasets: the MPI-Sintel, the Middlebury and the KITTI flow 2015 datasets. The Middlebury is represented only by six sequences (Grove2, Grove3, Hydrangea, RubberWhale, Urban2 and Urban3) because other sequences do not provide the ground truth or only two frames are available.

We prepare Fig. 4 to demonstrate the advantage of our approach for visual comparison with the standard two-frame approach. Occluded regions are displayed with shadow in the optical flow ground truth image. One can see that our algorithm is able to estimate flow values in occluded regions, while the two-frame method produces noisy flow values in these regions. For example, the regions in the red bounding box in Fig. 4 illustrate our claim. Also our method is more accurate in non-occluded regions (the regions in the blue bounding box). It is important, because non-occluded regions expand their flow values over image boundaries (the regions in green bounding box).

In this subsection we also perform an experiment to obtain the final dense optical flow without backward flow check and interpolation. In this case, the cost volume pre-processing based on the bilateral filtering is added, like it is done in the paper [53]. Fig. 6 illustrates this experiment and shows that the filtering operation in the cost volume can improve the discrete optical flow estimation. Table 3 gives the quantitative evaluation results. One can see that for the MPI-Sintel dataset the accuracy of our estimation without the backward flow consistency check increases up to 50% in occluded regions and up to 9% in non-occluded regions in comparision with the two-frame approach, which are the original results of DCFlow algorithm before consistency check. In the case of the cost volume pre-filtering accuracy of the final result (three-frame+ in Table 3) increases further by 21% and 12% in non-occluded and occluded regions respectively.

In this experiment we also leverage a variational energy minimization post-processing method [49] to obtain our final

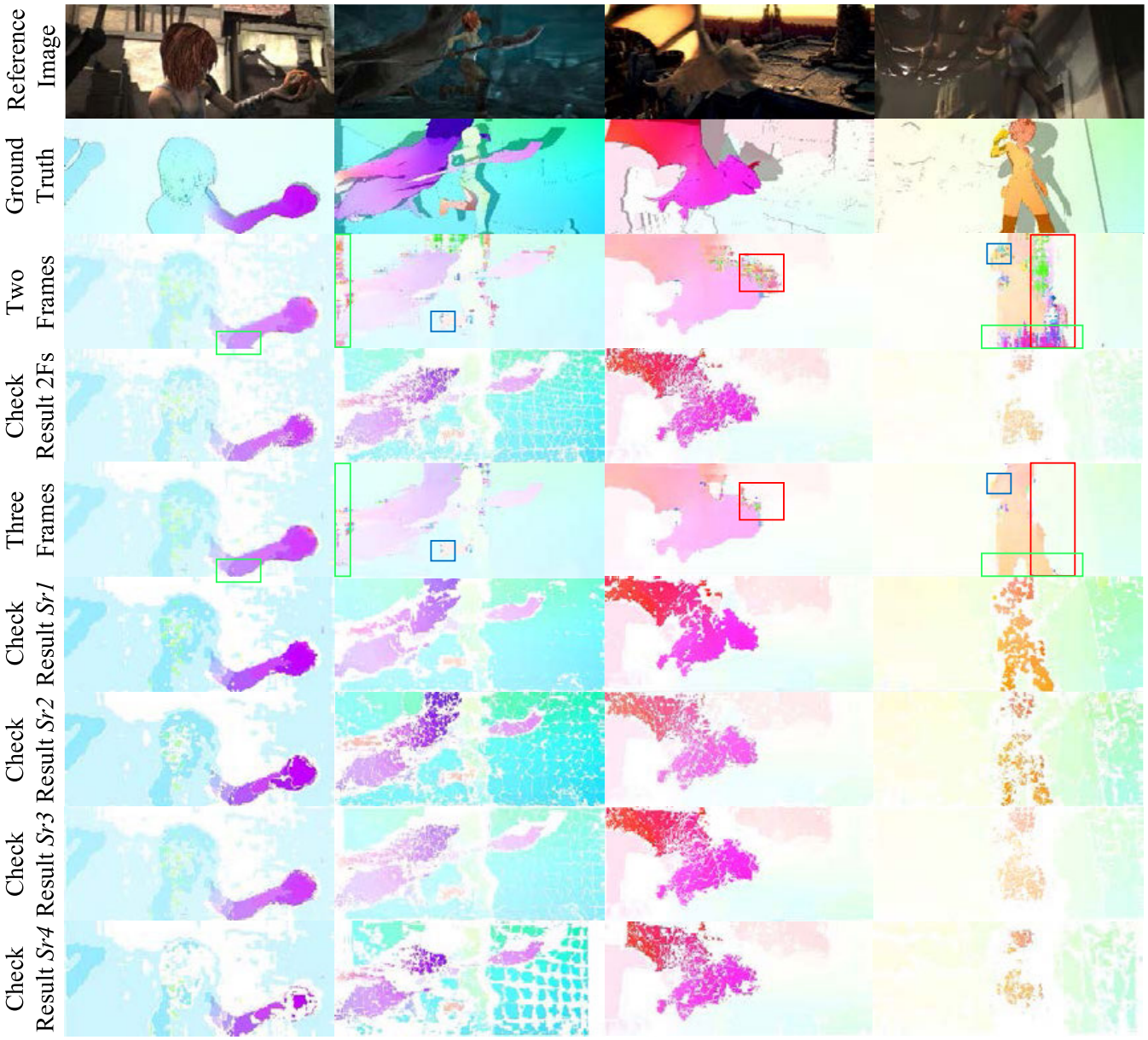


FIGURE 4. Optical flow results illustration: the first row illustrates the reference image; ground truth with occlusion mask (shadow area) is shown in the second row; the third row illustrates discrete optical flow using two frames; the two-frame approach results after consistency check is shown in the forth row; the fifth row illustrates the optical flow results with three frames; three-frame approach results with different outlier removal strategies are shown in the sixth - ninth rows.

optical flow results (three-frame++ in Table 3). In comparison with the DCFlow results after interpolation and post processing (Table 5), our method reaches the same accuracy level in non-occluded regions directly without the consistency check and interpolation. The result of this experiment demonstrates that potentially one can use our approach without the consistency check and interpolation parts.

C. RESULTS AFTER OUTLIERS HANDLING

In the previous subsection it is shown that accuracy of the optical flow estimation with our triple image matching cost formation is considerably higher than with the standard two-image matching cost. In this part, we test the impact of different outlier handling methods for the final optical flow estimation results (Step 3 in Fig. 3). The problem is that the popular consistency check procedure usually removes estimated flow values in occluded regions, thus we cannot capitalize advantages of the flow estimation that are achieved in the previous algorithmic step.

In this subsection we apply several different strategies for outlier handling, which include outlier removal based on the flow field map segmentation and several modifications of the consistency check procedure. These results are summarized

FIGURE 5. Proposed additional algorithm based on tripe image matching cost and cost volume filter: outlier handing and interpolation part are not necessary in this simplified pipeline.

FIGURE 6. Discrete optical flow results on the MPI-Sintel training dataset for two frames, three frames and three frames with cost volume filtering; final flow values after post-processing.

TABLE 4. Endpoint error and density of outliers handling results on the final pass of the MPI-Sintel dataset: DCFlow (sparse matching points) and four different outliers handling strategies.

Method	all	noc	occ	dens-all(%)	dens-occ(%)
DCFlow(smp)	**2.3808**	**2.1132**	9.4510	74.13	41.22
TIMCflow *Str*1	3.2187	2.4914	10.1091	**84.82**	**66.65**
TIMCflow *Str*2	3.4647	2.7555	**9.1800**	79.54	59.26
TIMCflow *Str*3	3.1035	2.6266	11.6307	81.31	52.37
TIMCflow *Str*4	2.6430	2.3106	10.1220	71.28	40.68

in Table 4. For Strategy1, we use the breadth-first search [54] technique to segment the flow field map and remove the regions with less than 20 pixels. For Strategy2, we estimate two different discrete flows relative to the same current frame f^t, but with different sets of λ. We use standard forward and backward flow consistency check to Strategy3, but with two different thresholds T_1 and T_2 for consistency check: T_1 is equal to 0.8 for area in which $C^{fk} < C^{bk}$ and T_2 is equal to 3 elsewhere. For Strategy4, we use backward flow computed also with three images but shift one frame compared with forward flow. Different outlier handling results can be seen in Fig. 4. Formally, the best results among our strategies is achieved with Strategy2 and Strategy4. However, these strategies produce a more sparse output, thus making the final optical flow estimation worse than the output of Strategy3.

D. DENSE OPTICAL FLOW RESULTS AFTER INTERPOLATION AND REFINEMENT

We consider two state-of-the-art interpolation methods in our experiments: EpicFlow [14] and InterpoNet [48] to get dense initialization flow values for the final variational refinement. The default parameters of EpicFlow and InterpoNet are the same for different outlier handling strategies. In the interpolation part of Table 5 one can see that our approach gets the best interpolation results with Strategy3 (circle 4 in Fig. 3), even for the refinement results (circle 5 in Fig. 3). We find that the interpolation result of the InterpoNet method is better than the result obtained with the EpicFlow algorithm, especially for occluded regions.

TABLE 5. Interpolation and variational refinement results: quantitative comparison of interpolation and refinement results with DCFlow (sparse matching points) and for four different outliers handling strategies.

	EpicFlow			InterpoNet		
Interpolation	all	noc	occ	all	noc	occ
with DCFlow(smp)	4.71	3.48	11.17	4.60	3.42	10.13
TIMCflow *Str*1	5.19	3.58	12.97	4.54	3.40	9.73
TIMCflow *Str*2	5.32	3.85	12.48	4.82	3.70	9.79
TIMCflow *Str*3	**4.55**	**3.32**	**10.82**	**4.41**	**3.22**	**9.64**
TIMCflow *Str*4	5.32	3.92	12.25	5.10	3.37	10.81
Refinement	all	noc	occ	all	noc	occ
with DCFlow(smp)	4.13	2.91	10.68	4.01	2.81	9.78
TIMCflow *Str*1	4.46	2.92	12.02	3.92	2.78	9.34
TIMCflow *Str*2	4.59	3.15	11.62	4.40	3.27	9.53
TIMCflow *Str*3	**3.99**	**2.73**	**10.61**	**3.82**	**2.61**	**9.30**
TIMCflow *Str*4	4.69	3.28	11.87	4.44	3.06	10.39

E. ADDITIONAL ALGORITHM

The intermediate results of our algorithm considerably outperform the DCFlow method, however the final performance gain is not that significant, we propose an additional algorithm in Fig. 5 that simplifies the proposed calculation scheme by removing the outlier removal and interpolation steps of the original pipeline. Consequently, this innovation decreases computation complexity. In this case, the results are not better, but comparable with the original results of

TABLE 6. Running time of different methods (sec).

Method	Cost Volume	Optimization	Interpolation	refinement	total
Two-Frame	0.35	2.50	0.41	1	4.24
Three-Frame	0.43	2.63	0.41	1	4.47
Three-Frame_add	0.87	1.19	–	1	3.06

the two-frame pipeline, however, this decreases the computational complexity of the algorithm.

F. RUNNING TIME

We report and compare running time of our main pipeline (in Fig. 3 and additional algorithm (in Fig. 5) in Table 6. Our three-frame version increases the calculation time minimally. In contrast, our additional algorithm Three-Frame_add algorithm considerably decreases the computational time.

VI. CONCLUSION

In this paper, we propose a new matching cost formation based on two assumptions: most occlusion regions that are invisible in the forward frame image (relative to the current frame) are visible in the backward frame; the forward flow is approximately equal to the negative value of the backward flow. The assumptions allow us to form the composite matching cost as a combination of two independent forward and backward matching costs. The proposed method allows us to improve the standard two-frame matching technique. Consequently, our approach considerably increases discrete optical flow estimation after the matching cost processing. Experimental results have shown that our TIMCflow pipeline can get better results than two-frame pipeline and reach three first rank positions among nine metrics. In addition, we also propose a simplified pipeline without consistency check and interpolation that can keep comparable accuracy. The running time of our TIMCflow stays at the same level as the two-frame pipeline, and our reduced pipeline shortens running time significantly when compared to the full pipeline. Note that the consistency check procedure usually removes estimated flow values in occluded regions, thus we cannot fully capitalize advantages of the flow estimation that are achieved on the previous algorithmic step of our pipeline, and we plan to improve this aspect of our algorithm in future work.

ACKNOWLEDGMENT

The authors would like to acknowledge the CERCA Programme of Generalitat de Catalunya and also the generous GPU support from Nvidia.

REFERENCES

[1] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, nos. 1–3, pp. 185–203, Aug. 1981.

[2] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Artif. Intell.*, vol. 2, 1981, pp. 674–679.

[3] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. V. D. Smagt, D. Cremers, and T. Brox, "FlowNet: Learning optical flow with convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2758–2766.

[4] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "FlowNet 2.0: Evolution of optical flow estimation with deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jul. 2017, p. 6.

[5] A. Ranjan and M. J. Black, "Optical flow estimation using a spatial pyramid network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jul. 2017, p. 2.

[6] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8934–8943.

[7] Y. Lecun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989.

[8] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "Models matter, so does training: An empirical study of CNNs for optical flow estimation," 2018, *arXiv:1809.05571*. [Online]. Available: https://arxiv.org/abs/1809.05571

[9] C. Bailer, B. Taetz, and D. Stricker, "Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 4015–4023.

[10] C. Bailer, K. Varanasi, and D. Stricker, "CNN-based patch matching for optical flow with thresholded hinge embedding loss," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, vol. 2, no. 3, p. 5.

[11] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "DeepFlow: Large displacement optical flow with deep matching," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1385–1392.

[12] Y. Hu, R. Song, and Y. Li, "Efficient coarse-to-fine patchmatch for large displacement optical flow," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 5704–5712.

[13] T. Brox and J. Malik, "Large displacement optical flow: Descriptor matching in variational motion estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 500–513, Mar. 2011.

[14] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "EpicFlow: Edge-preserving interpolation of correspondences for optical flow," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1164–1172.

[15] Y. Hu, Y. Li, and R. Song, "Robust interpolation of correspondences for large displacement optical flow," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 481–489.

[16] B. Glocker, N. Paragios, N. Komodakis, G. Tziritas, and N. Navab, "Optical flow estimation with uncertainties through dynamic MRFs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8

[17] L. Xu, J. Jia, and Y. Matsushita, "Motion detail preserving optical flow estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1744–1757, Sep. 2012.

[18] M. G. Mozerov, "Constrained optical flow estimation as a matching problem," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 2044–2055, May 2013.

[19] M. Menze, C. Heipke, and A. Geiger, "Discrete optimization for optical flow," in *Proc. German Conf. Pattern Recognit.* Cham, Switzerland: Springer, 2015, pp. 16–28.

[20] Q. Chen and V. Koltun, "Full flow: Optical flow estimation by global optimization over regular grids," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4706–4714.

[21] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[22] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proc. ECCV*, 1994, pp. 151–158.

[23] J. P. Lewis, "Fast template matching," in *Proc. Vis. Interface, Can. Image Process. Pattern Recognit. Soc.*, Quebec City, QC, Canada, May 1995, pp. 120–123.

[24] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang, "On building an accurate stereo matching system on graphics hardware," in *Proc. GPUCV*, 2011, pp. 467–474.

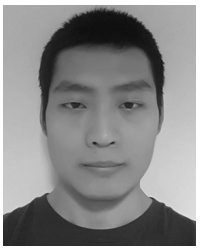
[25] D. Kong and H. Tao, "A method for learning matching errors for stereo computation," in *Proc. BMVC*, vol. 1, 2004, p. 2.

[26] M. Brown, G. Hua, and S. Winder, "Discriminative learning of local image descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 1, pp. 43–57, Jan. 2011.

[27] X. Han, T. Leung, Y. Jia, R. Sukthankar, and A. C. Berg, "MatchNet: Unifying feature and metric learning for patch-bsed matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3279–3286.

[28] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4353–4361.

[29] J. Zbontar and Y. LeCun, "Stereo matching by training a convolutional neural network to compare image patches," *J. Mach. Learn. Res.*, vol. 17, nos. 1–32, p. 2, 2016.
[30] J. Xu, R. Ranftl, and V. Koltun, "Accurate optical flow via direct cost volume processing," in *Proc. CVPR*, 2017, pp. 1289–1297.
[31] T. Brox, A. Bruhn, and J. Weickert, "Variational motion segmentation with level sets," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2006, pp. 471–483.
[32] D. W. Murray and B. F. Buxton, "Scene segmentation from visual motion using global optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vols. PAMI-9, no. 2, pp. 220–228, Mar. 1987.
[33] J. Weickert and C. Schnörr, "Variational optic flow computation with a spatio-temporal smoothness constraint," *J. Math. Imag. Vis.*, vol. 14, no. 3, pp. 245–255, May 2001.
[34] R. Kennedy and C. J. Taylor, "Optical flow with geometric occlusion estimation and fusion of multiple frames," in *Proc. Int. Workshop Energy Minimization Methods Comput. Vis. Pattern Recognit.* Cham, Switzerland: Springer, 2015, pp. 364–377.
[35] A. Salgado and J. Sánchez, "Temporal constraints in large optical flow estimation," in *Proc. Int. Conf. Comput. Aided Syst. Theory*. Berlin, Germany: Springer, 2007, pp. 709–716.
[36] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof, "Anisotropic Huber-L1 optical flow," in *Proc. BMVC*, vol. 1, no. 2, 2009, p. 3.
[37] S. Volz, A. Bruhn, L. Valgaerts, and H. Zimmer, "Modeling temporal coherence for optical flow," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1116–1123.
[38] M. Black and P. Anandan, "Robust dynamic motion estimation over time," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Dec. 2002, pp. 296–302.
[39] D. Maurer and A. Bruhn, "ProFlow: Learning to predict optical flow," 2018, *arXiv:1806.00800*. [Online]. Available: https://arxiv.org/abs/1806.00800
[40] C. Ballester, L. Garrido, V. Lazcano, and V. Caselles, "A TV-L^1 optical flow method with occlusion detection," in *Proc. Joint DAGM German Assoc. Pattern Recognit. OAGM Symp.* Berlin, Germany: Springer, 2012, pp. 31–40.
[41] J. Janai, F. Güney, A. Ranjan, M. Black, and A. Geiger, "Unsupervised learning of multi-frame optical flow with occlusions," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 690–706.
[42] M. Neoral, J. Šochman, and J. Matas, "Continual occlusions and optical flow estimation," 2018, *arXiv:1811.01602*. [Online]. Available: https://arxiv.org/abs/1811.01602
[43] Z. Ren, O. Gallo, D. Sun, M.-H. Yang, E. B. Sudderth, and J. Kautz, "A fusion approach for multi-frame optical flow estimation," 2018, *arXiv:1810.10066*. [Online]. Available: https://arxiv.org/abs/1810.10066
[44] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, in Lecture Notes in Computer Science, vol. 7577, A. Fitzgibbon, Eds. Berlin, Germany: Springer-Verlag, Oct. 2012, pp. 611–625.
[45] A. Ihler, J. Fisher, and A. Willsky, "Loopy belief propagation: Convergence and effects of message errors," *J. Mach. Learn. Res.*, vol. 6, pp. 905–936, May 2005.
[46] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
[47] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jul. 2005, pp. 807–814.
[48] S. Zweig and L. Wolf, "InterpoNet, a brain inspired neural network for optical flow dense interpolation," 2016, *arXiv:1611.09803*. [Online]. Available: https://arxiv.org/abs/1611.09803
[49] H. Zimmer, A. Bruhn, and J. Weickert, "Optic flow in harmony," *Int. J. Comput. Vis.*, vol. 93, no. 3, pp. 368–388, Jul. 2011.
[50] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume," 2017, *arXiv:1709.02371*. [Online]. Available: https://arxiv.org/abs/1709.02371
[51] M. Menze, C. Heipke, and A. Geiger, "Joint 3D estimation of vehicles and scene flow," in *Proc. ISPRS Workshop Image Sequence Anal. (ISA)*, 2015, pp. 427–434.
[52] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *Int. J. Comput. Vis.*, vol. 92, no. 1, pp. 1–31, 2011.
[53] M. G. Mozerov and J. Van De Weijer, "Accurate stereo matching by two-step energy minimization," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1153–1163, Mar. 2015.
[54] J. Silvela and J. Portillo, "Breadth-first search and its application to image processing problems," *IEEE Trans. Image Process.*, vol. 10, no. 8, pp. 1194–1199, Aug. 2001.



FEI YANG received the B.S. degree from Xidian University, Xi'an, China, in 2013, and the M.S. degree from Northwestern Polytechnical University, Xi'an, in 2016. He is currently pursuing the Ph.D. degree with the Department of Automation, Northwestern Polytechnical University, China, and with the Computer Vision Centre, Autonomous University of Barcelona, Spain. His research interests are in the areas of optical flow and deep learning.

YONGMEI CHENG received the M.Sc. and Ph.D. degrees from Northwestern Polytechnical University, in 1997 and 2001, respectively. She is currently a Professor with the Department of Automation, Northwestern Polytechnical University, Xi'an, China. Her research interests include signal and information processing, information fusion, tracking, and artificial intelligence.

JOOST VAN DE WEIJER received the Ph.D. degree from the University of Amsterdam, in 2005. From 2005 to 2007, he was a Marie Curie Intra-European Fellow of the LEAR Team, INRIA, Rhóne-Alpes, France. He is currently the Leader of the LAMP Team, Computer Vision Center, Universitat Autónoma de Barcelona. His main research interest includes machine learning applied to computer vision.

MIKHAIL G. MOZEROV (Member, IEEE) received the M.S. degree in physics from Moscow State University, in 1982, and the Ph.D. degree in digital image processing from the Institute of Information Transmission Problems, Moscow, in 1995. From 2006 to 2010, he was a Ramon y Cajal Fellow of the Universitat Autónoma de Barcelona, Barcelona, where he is currently a Senior Scientist with the Computer Vision Center, Department of Informatics. His research interests include signal and image processing, stereo and optical flow, and pattern recognition.