

# FLOWFIELDS++: ACCURATE OPTICAL FLOW CORRESPONDENCES MEET ROBUST INTERPOLATION

*René Schuster, Christian Bailer, Oliver Wasenmüller, Didier Stricker*

DFKI – German Research Center for Artificial Intelligence  
firstname.lastname@dfki.de

## ABSTRACT

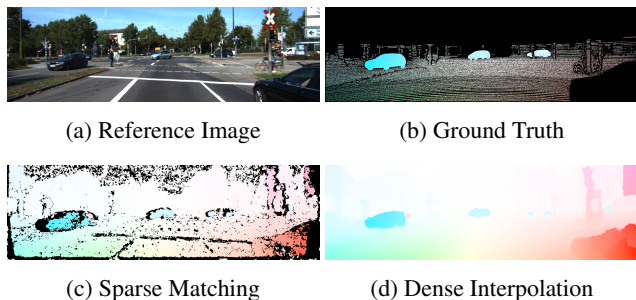
Optical Flow algorithms are of high importance for many applications. Recently, the Flow Field algorithm and its modifications have shown remarkable results, as they have been evaluated with top accuracy on different data sets. In our analysis of the algorithm we have found that it produces accurate sparse matches, but there is room for improvement in the interpolation. Thus, we propose in this paper FlowFields++, where we combine the accurate matches of Flow Fields with a robust interpolation. In addition, we propose improved variational optimization as post-processing. Our new algorithm is evaluated on the challenging KITTI and MPI Sintel data sets with public top results on both benchmarks.

**Index Terms**— Interpolation, KITTI, Matching, MPI Sintel, Optical Flow

## 1. INTRODUCTION

One core component of machine vision in many domains is the estimation of dense optical flow, e.g. as input for other algorithms such as 3D reconstruction or odometry, in autonomous driving and robot navigation as perception of the motion of the environment, and many more. More and more vision-based applications require increasing performance of their underlying algorithms. That includes faster run time to enable real time applications, higher accuracy to distinguish from competitors, or increased robustness under challenging environmental conditions to improve the reliability. The need for steadily improving optical flow estimation has led to impressive results in research and industry regarding the performance and diversity of approaches.

In this work, we push the limits further by creating an enhanced optical flow algorithm that performs superior on different data sets and is not dedicated to a single data set only. Towards this end, we have identified Flow Fields [1] as a very versatile state-of-the-art matching approach that we combine with robust interpolation (cf. Figure 1). The advantages of each separate concept shall surpass the weaknesses of their respective counterpart to boost the overall performance way beyond each individual algorithm. In detail, our contributions are the following:



**Fig. 1:** FlowFields++: Accurate matches get robustly interpolated for dense optical flow.

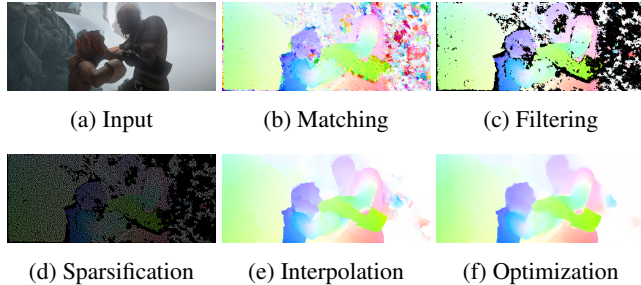
- Novel combination of accurate matching with robust interpolation.
- Improved variational optimization for optical flow in a dual-frame setting.
- Thorough evaluation<sup>1</sup> on two different challenging data sets to verify versatility and accuracy.

## 2. RELATED WORK

Flow Fields [1] can be considered as the basis of our method. It was among the first approaches to achieve top performance across multiple data sets and has been refined several times since its publication. Flow Fields+ [2] improved the algorithm by more sophisticated matching, FlowFieldsCNN [3] used deep learning to evaluate the matching cost, and ProbFlowFields [4] improved the results by jointly estimating optical flow and a certainty measure. Our approach shares the basic concepts with the mentioned methods. That is, we also perform dense matching by multi-scale propagation and random search, followed by outlier rejection, post processed by an interpolation mechanism. However, the novel FlowFields++ differs from those approaches by combining the matching accuracy of Flow Fields with robust interpolation.

Dense interpolation has become a very popular post processing step for many applications ever since the publication of the first successful interpolation method EPICFlow [5]. It

<sup>1</sup><https://youtu.be/fgmrTlJJRe4>



**Fig. 2:** Visualization of the FlowFields++ pipeline.

was used by Flow Fields and many other matching methods to produce dense results. InterpNet [6] tried to solve the task of optical flow interpolation with a neural network that showed improvements over EPICFlow depending on input matches and data set. Independent of these factors are the advantages of RICFlow [7] over EPICFlow. The basic idea of edge preserving interpolation is complemented by increased robustness in the computation of the piece-wise interpolation models. We will exploit this robustness for our approach and extend it further by improved edge detectors and adjusted variational refinement.

In recent eras of deep learning, there are also approaches that use convolutional neural networks (CNNs) to aid or solve the task of optical flow estimation. Some try to compute optical flow in an end-to-end manner [8, 9, 10], others use neural networks to compute a semantic segmentation as additional input [11, 12, 13], and some use deep learning for matching cost computation [3, 14, 15]. Of course, all deep learning approaches require a lot of proper training data and none yet has showed to generalize well across different data sets without retraining or tuning. Our approach maintains its versatility by avoiding deep learning.

### 3. METHOD

As mentioned before, our method consists mainly of two parts. First, the computation of sparse, accurate optical flow correspondences, and secondly, the robust interpolation to a dense flow field (see Figure 1). As others before, we apply a coarse-to-fine matching approach using spatial propagation and random search [1]. A strong multi-stage filtering step is reliably removing most of the falsely matched correspondences. Robust, edge-aware interpolation fills up the filtered gaps efficiently [7]. Sharp edges are obtained via a random forest trained on semantic boundaries [16, 17]. Final refinement is done via an adjusted variational framework [18]. The separate steps are visualized in Figure 2.

#### 3.1. Matches

We define the matching cost of corresponding pixels as a distance between their feature descriptors. For our experiments

we consider two terms. Either, the Hamming distance of the binary patterns created by the Census transform [19] in a  $7 \times 7$  patch for all channels of the CIELab color space, or the Euclidean distance of SIFT features [20]. Sub-pixel locations are interpolated bilinearly in both cases.

For initialization on lowest resolution, Walsh-Hadamard feature descriptors are matched using kD-tress [21]. Those initial matches are propagated into all four quadrants for  $i = 12$  iterations at each scale. We use sub-scales and sub-sub-scales as in [2]. Additionally, after each of the last three propagation steps of one iteration, a random search is performed where a small random offset is added to each flow component. To reject outliers introduced during the matching process, we perform two consistency checks and region-based filtering during sparsification. Each consistency check compares the estimated optical flow with an inverse flow field computed with different matching parameters and removes the estimated flow if the consistency error exceeds a consistency threshold  $\epsilon$ . Finally after filtering, we sparsify the remaining matches further by only selecting matches with the lowest consistency error in each  $3 \times 3$  block that contains at least  $s$  matches. The sparsification helps to increase the spatial support during interpolation and has a positive effect on the run time. Useful implementation details are given in [2].

#### 3.2. Interpolation

Our interpolation strategy has two core properties: Edge-awareness and robustness. Following the ideas of [5, 7] we compute local neighborhoods with respect to a geodesic distance that is based on image edges. To simulate euclidean distances, an offset of 0.002 is added to the edge maps. Contrary to the original work in this field, we do not rely on basic image edges but rather train a random forest on semantic boundaries using the approach of [17]. It was recently shown that this kind of edges are superior to the standard SED [16] model. Further we use the approximation of [7] by segmenting the images into super-pixels to find an optimal consensus set out of a local super-pixel neighborhood to compute the interpolation model. Since random sampling alone would be too inefficient for model estimation, this approach performs spatial propagation of already estimated models. The interpolation model is an affine 2D transformation of 6 unknowns that is used to transform all pixels where optical flow values are missing. Compared to previous methods like EPICFlow [5], this strategy is much more robust because model estimation is not based on all matches of the local neighborhood. This way, the additional inlier selection can successfully reject remaining outliers in the matches.

#### 3.3. Variational Optimization

The densely interpolated flow field is then used as initialization for variational refinement. We use the framework of Brox

**Table 1: Results on KITTI 2015.** We compare average percentage of outliers on background *Fl-bg*, foreground *Fl-fg*, and all *Fl-all* pixels. The best published monocular methods are listed. Multi-frame and deep learning approaches are given in gray. Run times on a GPU are in parenthesis.

Method	Fl-bg	Fl-fg	Fl-all	Time
MirrorFlow [22]	8.93	17.07	10.29	660 s
SDF [12]	8.61	23.01	11.01	—
UnFlow [10]	10.15	15.93	11.11	(0.12 s)
CNNF+PMBP [23]	10.08	18.56	11.49	2700 s
MR-Flow [24]	10.13	22.51	12.19	480 s
DCFlow [25]	13.10	23.70	14.86	(8.6 s)
<b>FlowFields++ (ours)</b>	<b>14.82</b>	<b>17.77</b>	<b>15.31</b>	<b>29 s</b>
SOF [11]	14.63	22.83	15.99	360 s
JFS [13]	15.90	19.31	16.47	780 s
DF+OIR [26]	15.11	23.45	16.50	180 s
ImpPB+SPCI [27]	17.25	20.44	17.78	(60 s)
FlowFieldCNN [3]	18.33	20.42	18.68	(23 s)
RicFlow [7]	18.73	19.09	18.79	5 s
FlowFields+ [2]	19.51	21.26	19.80	28 s

et al. [18] with two important adjustments. First, the coarse-to-fine pyramid steps are not required because our dense flow field is already a very good estimate on full resolution. Secondly, we do not optimize the optical flow where it would leave the image domain. This is very important because the variational energy at those regions is solely defined by the smoothness term. Thus, the optimization process tends towards constant optical flow which is often inappropriate. By leaving these areas out, we rather rely on our already precise interpolation instead of taking the high risk of oversmoothing. Evidence for this decision is given by the out-of-bounds regions of the KITTI data set in Figure 3.

## 4. RESULTS

We evaluate our method on the popular optical flow benchmarks KITTI [29] and MPI Sintel [30] and show that we belong to the few methods that achieve top performance on both. This again confirms that our generic approach is not restricted to any data domain or setup.

### 4.1. Parameter Selection

Unless explicitly stated here, we use the same values for both data sets. As matching features, we use SIFT for KITTI and Census for Sintel as described in Section 3.1. We use different random forests, i.e on KITTI we estimate semantic boundaries as described in Section 3.2 and on Sintel we use the default model of [16]. Other variations are given in Table 3. All other parameters apply to both data sets and are either given in the previous sections or are the default values of the respective framework.

**Table 2: Results on MPI Sintel.** We give the average end-point error (EPE) on the *final* rendering pass for *matched* and *unmatched* regions and *all* pixels. Only the best variant of each published method is listed. Multi-frame and deep learning approaches are given in gray.

Method	EPE all	EPE matched	EPE unmatched
DCFlow [25]	5.119	2.283	28.228
FlowFieldsCNN [3]	5.363	2.303	30.313
MR-Flow [24]	5.376	2.818	26.235
S2F-IF [28]	5.417	2.549	28.795
<b>FlowFields++ (ours)</b>	<b>5.486</b>	<b>2.614</b>	<b>28.900</b>
InterpoNet [6]	5.535	2.372	31.296
RicFlow [7]	5.620	2.765	28.907
ProbFlowFields [4]	5.696	2.545	31.371
FlowFields+ [2]	5.707	2.684	30.642
DeepDiscreteFlow [15]	5.728	2.623	31.042
FlowNet2 [9]	5.739	2.752	30.108
FlowFields [1]	5.810	2.621	31.799

**Table 3: Parameters for KITTI and MPI Sintel.**

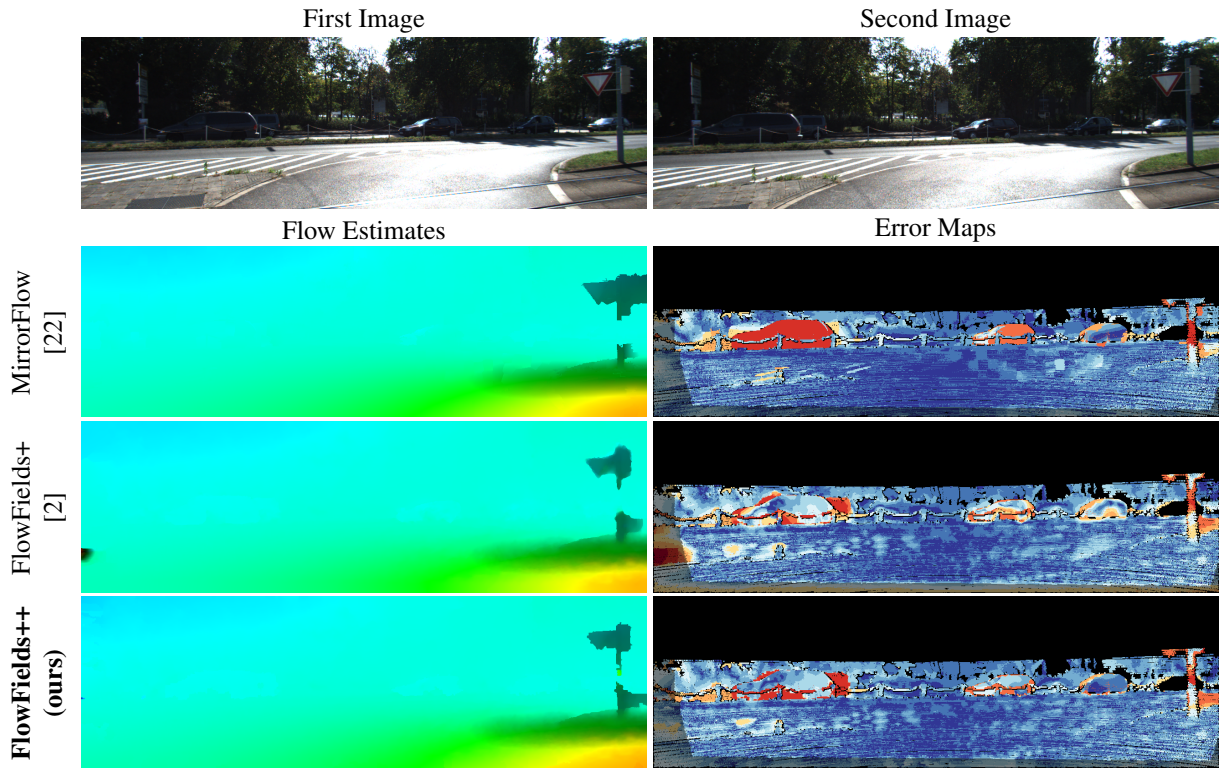
Parameter	KITTI	Sintel
Consistency threshold $\epsilon$	1	7
Minimum matches $s$	7	4
Super-pixel size	20	50
Local neighborhood size	150	200
Variational iterations	2	5

### 4.2. KITTI 2015

The first data set on which we evaluate our approach is KITTI [29]. It consists of traffic scenarios and provides sparse ground truth. The data set contains large displacements and challenging lightning conditions.

For a visual impression, we give exemplary results of our method compared to others in Figure 3. Worth highlighting is the impact of our adjusted variational refinement described in Section 3.3. Compared to Flow Fields+ that uses the variational optimization of EPICFlow, our approach estimates the out-of-bounds regions correctly. This can be seen in the shaded regions of the error maps in Figure 3.

For a qualitative comparison, we show the top performing methods on KITTI in Table 1. Only three non-learning-based methods perform better than our approach. However, MirrorFlow [22] is not competitive on other data sets, MR-Flow [24] is using more than two frames which resolves ambiguities in invisible parts of the reference image, and DCFlow [25] performs much worse on moving foreground regions that are of high importance in motion estimation. Compared to Flow Fields+ [2] that shares our matching strategy and to RICFlow [7] that also uses robust interpolation, we could reduce the average amount of outliers by about 20 %.



**Fig. 3:** Visual results on the test data of KITTI. We show the estimated optical flow and the according error maps for MirrorFlow, FlowFields+, and our FlowFields++. Contrary to FlowFields+, our approach estimates out-of-bounds regions correctly.

### 4.3. MPI Sintel

In addition, we use MPI Sintel [30] for evaluation. This data set provides dense ground truth for synthetically rendered sequences. It typically consists of large displacements with fine details, highly non-rigid motions of close-up deformable characters, and challenging image quality due to motion blur and atmospheric effects in the *final* rendering pass (cf. Figure 2). This characteristics are often the reason why algorithms that perform good on KITTI tend to perform much worse on Sintel. Our FlowFields++ belongs to the rare exceptions along with DCFlow [25] and MR-Flow [24]. As mentioned before MR-Flow is using multiple frames which is the reason for its exceptional performance in *unmatched* regions, but in *matched* areas its performance is considerably worse than all other listed methods.

Table 2 lists the highest ranked methods on Sintel. Noticeably few methods are also listed in the KITTI ranking. Especially deep learning approaches are rarely represented with top performance on both data sets, besides FlowFieldsCNN [3]. Again, only three non-learning approaches are ranked higher than FlowFields++ while only one of those is a dual-frame method that is also represented on KITTI. On the *clean* rendering pass FlowFields++ is even the highest ranked dual-frame method.

## 5. CONCLUSION

We have presented FlowFields++, a generic optical flow algorithm by combining highly accurate matching and robust, edge-preserving interpolation. The performance of the approach was evaluated on two diverse public data sets. In a joint ranking of both data sets, we claim to be the second best method among all dual-frame methods (even including deep learning approaches) after DCFlow [25] that we still outperform in foreground regions on KITTI.

## 6. REFERENCES

- [1] Christian Bailer, Bertram Taetz, and Didier Stricker, “Flow Fields: Dense correspondence fields for highly accurate large displacement optical flow estimation,” in *International Conference on Computer Vision (ICCV)*, 2015.
- [2] Christian Bailer, Bertram Taetz, and Didier Stricker, “Optical Flow Fields: Dense correspondence fields for highly accurate large displacement optical flow estimation,” *arXiv preprint arXiv:1703.02563*, 2017.
- [3] Christian Bailer, Kiran Varanasi, and Didier Stricker, “CNN-based patch matching for optical flow with thresholded hinge embedding loss,” in *Conference*

- on *Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [4] Anne S. Wannenwetsch, Margret Keuper, and Stefan Roth, “Probflow: Joint optical flow and uncertainty estimation,” in *International Conference on Computer Vision (ICCV)*, 2017.
- [5] Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, and Cordelia Schmid, “EpicFlow: Edge-preserving interpolation of correspondences for optical flow,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [6] Shay Zweig and Lior Wolf, “InterpoNet, a brain inspired neural network for optical flow dense interpolation,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [7] Yinlin Hu, Yunsong Li, and Rui Song, “Robust interpolation of correspondences for large displacement optical flow,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [8] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox, “Flownet: Learning optical flow with convolutional networks,” in *International Conference on Computer Vision (ICCV)*, 2015.
- [9] Eddy Ilg, Nikolaus Mayer, Tommo Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox, “FlowNet 2.0: Evolution of optical flow estimation with deep networks,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [10] Simon Meister, Junhwa Hur, and Stefan Roth, “UnFlow: Unsupervised learning of optical flow with a bidirectional census loss,” in *Conference on Artificial Intelligence (AAAI)*, 2018.
- [11] Laura Sevilla-Lara, Deqing Sun, Varun Jampani, and Michael J Black, “Optical flow with semantic segmentation and localized layers,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [12] Min Bai, Wenjie Luo, Kaustav Kundu, and Raquel Urtasun, “Exploiting semantic information and deep matching for optical flow,” in *European Conference on Computer Vision (ECCV)*, 2016.
- [13] Junhwa Hur and Stefan Roth, “Joint optical flow and temporally consistent semantic segmentation,” in *European Conference on Computer Vision (ECCV)*, 2016.
- [14] Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, and Cordelia Schmid, “DeepMatching: Hierarchical deformable dense matching,” *International Journal of Computer Vision (IJCV)*, 2016.
- [15] Fatma Güney and Andreas Geiger, “Deep discrete flow,” in *Asian Conference on Computer Vision (ACCV)*, 2016.
- [16] Piotr Dollár and C Lawrence Zitnick, “Structured forests for fast edge detection,” in *International Conference on Computer Vision (ICCV)*, 2013.
- [17] René Schuster, Oliver Wasenmüller, Georg Kuschik, Christian Bailer, and Didier Stricker, “SceneFlow-Fields: Dense interpolation of sparse scene flow correspondences,” in *Winter Conference on Applications of Computer Vision (WACV)*, 2018.
- [18] Thomas Brox, Andrés Bruhn, Nils Papenberg, and Joachim Weickert, “High accuracy optical flow estimation based on a theory for warping,” *European Conference on Computer Vision (ECCV)*, 2004.
- [19] Ramin Zabih and John Woodfill, “Non-parametric local transforms for computing visual correspondence,” in *European Conference on Computer Vision (ECCV)*, 1994.
- [20] David G Lowe, “Object recognition from local scale-invariant features,” in *International Conference on Computer Vision (ICCV)*, 1999.
- [21] Yacov Hel-Or and Hagit Hel-Or, “Real-time pattern matching using projection kernels,” *Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2005.
- [22] Junhwa Hur and Stefan Roth, “MirrorFlow: Exploiting symmetries in joint optical flow and occlusion estimation,” in *International Conference on Computer Vision (ICCV)*, 2017.
- [23] Feihu Zhang and Benjamin W Wah, “Fundamental principles on learning new features for effective dense matching,” *Transactions on Image Processing (IP)*, 2018.
- [24] Jonas Wulff, Laura Sevilla-Lara, and Michael J Black, “Optical flow in mostly rigid scenes,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [25] Jia Xu, Rene Ranftl, and Vladlen Koltun, “Accurate optical flow via direct cost volume processing,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [26] D. Maurer, M. Stoll, and A. Bruhn, “Order-adaptive and illumination aware variational optical flow refinement,” in *British Machine Vision Conference (BMVC)*, 2017.
- [27] Tal Schuster, Lior Wolf, and David Gadot, “Optical flow requires multiple strategies (but only one network),” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [28] Y. Yang and S. Soatto, “S2F: Slow-to-fast interpolator flow,” *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [29] Moritz Menze and Andreas Geiger, “Object scene flow for autonomous vehicles,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [30] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, “A naturalistic open source movie for optical flow evaluation,” in *European Conference on Computer Vision (ECCV)*, 2012.