MirrorFlow:
Exploiting Symmetries in Joint Optical Flow and Occlusion Estimation

Junhwa Hur Stefan Roth
Department of Computer Science, TU Darmstadt

Abstract

Optical flow estimation is one of the most studied problems in computer vision, yet recent benchmark datasets continue to reveal problem areas of today’s approaches. Occlusions have remained one of the key challenges. In this paper, we propose a symmetric optical flow method to address the well-known chicken-and-egg relation between optical flow and occlusions. In contrast to many state-of-the-art methods that consider occlusions as outliers, possibly filtered out during post-processing, we highlight the importance of joint occlusion reasoning in the optimization and show how to utilize occlusion as an important cue for estimating optical flow. The key feature of our model is to fully exploit the symmetry properties that characterize optical flow and occlusions in the two consecutive images. Specifically through utilizing forward-backward consistency and occlusion-disocclusion symmetry in the energy, our model jointly estimates optical flow in both forward and backward direction, as well as consistent occlusion maps in both views. We demonstrate significant performance benefits on standard benchmarks, especially from the occlusion-disocclusion symmetry. On the challenging KITTI dataset we report the most accurate two-frame results to date.

1. Introduction

Optical flow estimation has been studied intensely for several decades. Yet, recent optical flow benchmark datasets [8, 12] reveal challenges that current methods are still struggling to handle. Besides complex large motion and severe illumination changes, occlusions continue to pose a key challenge. Proper occlusion handling, especially in the presence of large motion, is becoming one of the critical factors to a method’s success in challenging scenes.

Occlusion estimation is a well-known chicken-and-egg problem that optical flow has been entangled with for a long time [1, 22, 30, 35]. Accurate knowledge of occluded areas is crucial for reliable optical flow estimation in order to prevent non-occluded areas from being adversely affected by occluded pixels. Yet, occlusion is a consequence of motion. Estimating accurate optical flow, conversely, is required for localizing occlusions reliably. We thus posit that their mutual dependency necessitates taking a joint approach and argue that this has not been done to the extent possible.

The majority of recent work instead addresses this challenging joint problem indirectly by considering occlusions as outliers of low-level correspondence estimation, e.g. [3, 9, 13, 17, 51]. Such approaches aim to mitigate the effects of occlusion by exploiting that occluded pixels gener-
ally violate the underlying model assumptions as there are no corresponding pixels in the other frame. Using a robust, truncated penalty in the data term naturally reduces the effects of the high data cost from occlusions, but also more generally from outlier pixels that violate brightness constancy [9, 29, 51]. Checking the forward-backward motion consistency in a subsequent post-processing step and extrapolating flow into inconsistent regions also helps resolving the motion mismatch in the occluded area [9, 11, 17, 29]. These simple procedures have been shown to diminish the influence of occlusions in practice.

Such strategies, however, still cannot completely free optical flow estimation from the ill effects of occlusion. Using a truncated data term leaves the possibility that occluded pixels can be incorrectly matched to other pixels for which the data cost is lower than the truncation constant. Additionally, when extrapolating flow in post-processing, false-positive matches may remain even after the forward-backward consistency check and can cause erroneous estimates to be propagated across a local region [31]. We thus argue that only the accurate localization of occluded regions, formulated as a joint estimation together with the flow, can fundamentally resolve this intertwined problem.

In this paper, we address the chicken-and-egg problem of optical flow and occlusion map estimation and propose a joint energy formulation and optimization method. Our approach directly utilizes their relationship and allows them to leverage each other through estimating forward and backward flow as well as occlusion maps for both directions all together. We exploit two key symmetry properties of the optical flow field and the occlusion map within the two consecutive images: forward-backward flow consistency and occlusion-disocclusion symmetry. These symmetry properties not only couple optical flow with occlusion, but also allow to exploit the geometric and temporal information in the two consecutive images to a greater extent.

The key contributions of our paper are as follows. To the best of our knowledge, we are the first to exploit the occlusion-disocclusion symmetry in joint optical flow and occlusion estimation and show its significant accuracy gains. Second, we demonstrate how this joint, symmetric treatment combined with a piecewise rigid formulation allows estimating optical flow without post-processing. Our experimental results demonstrate state-of-the-art accuracy on public benchmark datasets, where we improve the results especially in occluded areas. For the challenging KITTI dataset we report the most accurate results among two-frame methods to date, outperforming the latest approaches based on high-capacity deep networks [2, 10, 13, 19]. The fact that we are able to do so without employing learning demonstrates the significant benefits of our joint, symmetric flow and occlusion formulation.

2. Related Work

We can categorize occlusion handling approaches in optical flow roughly into two families of solutions.

**Occlusions as outliers.** As discussed above, the chicken-and-egg problem of flow and occlusion estimation becomes much simpler if occlusions are treated as outliers that violate the basic optical flow assumptions (e.g., brightness constancy and/or forward-backward flow consistency assumptions). A number of recent algorithms [3, 9, 11, 13, 17, 25, 29] follow a common strategy for outlier filtering. Based on a robust, truncated data term, they (i) separately estimate forward and backward flow with an asymmetric method, (ii) conduct a bi-directional consistency check, and (iii) interpolate flows into the outlier pixels in a post-processing stage, c.f. Fig. 2a. With the aid of highly capable interpolation methods [26, 31], this pipeline has been regarded as a well-justified practice. However, occasional failures during post-processing are inevitable and irreversible, and thus constitute a fundamental limitation.

In contrast, our symmetric approach explicitly integrates occlusions into the objective in order to exploit them as an important cue for the flow itself. Also, as shown in Fig. 2b, our integrative approach simultaneously estimates flow in both directions and thus encourages bi-directional consistency of the flow as part of the formulation, which naturally makes any post-processing unnecessary.

**Occlusions in a joint objective.** An outlier is a failure of flow estimation, but an occlusion is a consequence of motion, and can conversely be used as additional evidence for estimating optical flow. Distinguishing between the two opens new possibilities. A number of previous works consider occlusion explicitly in the formulation, but received less attention. We aim to bring them back into focus, revisit their ideas, and highlight the importance of jointly handling occlusions as a feature complementary to other recent trends, including the use of deep learning for appearance matching [2, 10, 13, 19].

All methods in this category begin with including an occlusion variable in their objective. Yet, they differ in the particular characteristics of occlusion utilized and how these are formulated. One basic way to characterize occlusion stems from the observation that brightness constancy mostly does not hold in occlusion areas due to the non-existence of corresponding pixels. Several approaches [18, 36, 41, 47] adopt a constant penalty (or truncated cost) in the data term so that it can (i) naturally lead to occluded pixels taking the constant penalty rather than a potentially higher matching cost, and (ii) explicitly exclude their matching cost from the objective (e.g., as visualized in Fig. (6) of [36]). However, this property alone is not sufficient [47] as it is impossible to discriminate between occluded pixels and pixels with strong illumination changes,
which both violate the assumption.

Introducing a forward-backward consistency constraint into the objective function is another useful strategy. Its benefit is that pixels are forced to be either visible and satisfy the bi-directional flow consistency, or are identified as occlusions [20]. This condition provides an additional cue for joint estimation. Yet, one should not forget that occlusion is a consequence of two different motions causing one pixel to geometrically occlude another. Layered optical flow models [36, 37] or 3D scene flow methods [42] can explicitly model the local depth relationship between layers and estimate occlusions simultaneously. Similarly, one can calculate overlapping areas between two triangular patches and detect occlusions by comparing the photometric cost [21]. Calculating the divergence of the motion field [5] or finding unique configurations of corresponding pixels [22, 41] can be alternative approaches.

Similar to ours, Alvarez et al. [1] exploit symmetry properties for jointly estimating optical flow and occlusion, except that their method does not utilize the occlusion and disocclusion symmetry. Interestingly, the most closely related work by Sun et al. [38] is from the stereo matching literature, except that forward-backward consistency is not used.

Though all these previous approaches successfully argue the effectiveness of their underlying models, each one omits at least one property that could be utilized. In contrast, our model exploits the relations and symmetry properties between optical flow and occlusion jointly and more completely, leading to significant benefits in accuracy.

3. Joint, Symmetric Approach

From two consecutive images, we jointly estimate optical flow maps in both directions and corresponding occlusion maps by fully exploiting their symmetries. The symmetry properties (i) couple the two different problem domains, (ii) better utilize the available image evidence, and (iii) lead to a well-balanced solution during optimization.

The first symmetry we consider is bi-directional motion consistency, i.e., motions of corresponding pixels that are visible in both views should be the inverse of one another. Unlike most previous work, we integrate this consistency into the energy, which not only leads to better estimates of both forward and backward flow through iterative optimization, but also obviates conventional post-processing.

Occlusion-disocclusion symmetry is the second property we consider, which geometrically explains how occlusions arise from differing motions of dynamic entities. As illustrated in the third and fourth column of Fig. 3, occlusions and disocclusions demonstrate a symmetry relationship, i.e., occlusions in the forward direction correspond to disocclusions in the backward direction and vice versa.

3.1. Piecewise rigid optical flow model

Our optical flow model is based on a piecewise rigid representation, which has recently been found to allow the effective regularization of 8-DoF or 9-DoF rigid motion of entities in images, yet still represent both diverse and general motions [15, 18, 28, 44, 51]. We first decompose an image into a set of superpixels [53] as shown in Fig. 3, and estimate the 8-DoF homography motion H of each superpixel. Each superpixel represents a possible surface in the scene, and the homography represents a locally rigid motion of the surface. Using this model also facilitates formulating the occlusion-disocclusion symmetry property in a comprehensive way, which will be explained below.

3.2. Joint energy with symmetries

Given the two consecutive images Iᵗ and Iᵗ⁺¹ with their superpixel representations, our model jointly estimates (i)
the forward motion $H^f$ (i.e., $I^t \rightarrow I^{t+1}$) and the backward motion $H^b$ (i.e., $I^{t+1} \rightarrow I^t$) of each superpixel, and (ii) the per-pixel occlusion maps $o^t$ and $o^{t+1}$ for each view. We formulate this through the energy

$$ E(H^f, H^b, o^t, o^{t+1}) = E_D(H^f, H^b, o^t, o^{t+1}) + \lambda_p E_P(H^f, H^b, o^t, o^{t+1}) + \lambda_C E_C(H^f, H^b, o^t, o^{t+1}) + \lambda_S E_S(o^t, o^{t+1}), $$

which consists of a data term $E_D$, a pairwise term $E_P$, a forward-backward consistency term $E_C$, and an occlusion-disocclusion symmetry term $E_S$.

### 3.2.1 Data term

The data term accumulates photometric differences across all pixels in both views given the 8-DoF homography motions of superpixels $H$ and the per-pixel occlusion masks $o_p$ in both views:

$$ E_D(H^f, H^b, o^t, o^{t+1}) = \sum_{p \in I^t} D^f_p + \sum_{p \in I^{t+1}} D^b_p $$

with

$$ D^f_p = \overline{o^t_p} \cdot \rho^f_D(p, H^f_s_p) + o^t_p \lambda_{occ} $$

$$ D^b_p = \overline{o^{t+1}_p} \cdot \rho^b_D(p, H^b_s_p) + o^{t+1}_p \lambda_{occ}. $$

For a non-occluded pixel $p$ (i.e., $\overline{o_p} = (1 - o_p) = 1$), the function $\rho_D(p, H^f_s)$ measures the truncated photometric error between pixel $p$ and its corresponding pixel $H^f_s_p$ in the other frame. $H^f_s$ is the homography motion of superpixel $s_p$ at pixel position $p$. We use a weighted sum of a gradient constancy term and a ternary transform [34], which is known to be robust under illumination changes [14, 43, 44]. When the corresponding location $H^b_s p$ falls outside the image boundary, $\rho_D(\cdot, \cdot)$ outputs the truncation constant $\tau_D$.

More specifically, we use a continuous version of the ternary transform, which can improve localization [43] compared to the conventional discrete setting. Furthermore, when calculating the ternary value in the other frame, we transform $7 \times 7$ patches from the reference frame to the other using the given homography $H^b_s$, and calculate the ternary transform based on the transformed patches. Using this strategy yields a more comprehensive data cost that is invariant to local shape deformation caused by the motion. We observe this to increase the flow accuracy; see supplementary material for details and a quantitative analysis.

For occluded pixels (i.e., $o_p = 1$), the constant penalty $\lambda_{occ}$ is applied so that we can avoid trivial cases in which all pixels are occluded or move outside the image boundary. We set $\lambda_{occ} < \tau_D$ so that pixels whose corresponding location is outside of the image boundary can be naturally inferred as occluded pixels during the optimization.

### 3.2.2 Pairwise term

The pairwise term penalizes the motion differences and occlusion status differences in an 8-neighborhood $N(p)$:

$$ E_P(H^f, H^b, o^t, o^{t+1}) = \sum_{p \in I^t} P^f_p + \sum_{p \in I^{t+1}} P^b_p $$

with

$$ P^f_p = \sum_{q \in N(p)} \left( \phi(H^f_s_p, H^f_s_q, p) + \lambda_O [o^t_q \neq o^t_{\bar{q}}] \right) $$

$$ P^b_p = \sum_{q \in N(p)} \left( \phi(H^b_s_p, H^b_s_q, p) + \lambda_O [o^{t+1}_q \neq o^{t+1}_{\bar{q}}] \right), $$

where $\lambda_O$ is a weight for the pairwise occlusion cost and $[\cdot]$ denotes the Iverson bracket. $\phi(H^b_s_p, H^b_s_q, p)$ measures
the difference of two motions induced by the homographies \(H_{sp}\) and \(H_{sq}\) at neighboring pixels \(p\) and \(q\), evaluated in the middle between the pixels \(\bar{p} = \frac{p + q}{2}\). This function distinguishes three different types of pairwise relationships \([18, 49, 50]\) in order to effectively express geometric relations that two neighboring superpixels can have (e.g., co-planar, hinge, others) in terms of their homography motion:

\[
\phi(H_{sp}, H_{sq}, p) = w_{p,q} \cdot \min(\phi_{\text{co}}, \phi_{h}, \tau_{p}) \tag{4a}
\]

with

\[
\phi_{\text{co}} = \frac{1}{|s_p \cup s_q|} \sum_{i \in s_p \cup s_q} \|H_{sp}p_i - H_{sq}p_i\| \tag{4b}
\]

\[
\phi_{h} = \|H_{sp}\bar{p} - H_{sq}\bar{p}\| + \lambda_{h}, \tag{4c}
\]

where \(\tau_{p}\) is a truncation constant and \(\lambda_{h}\) a constant bias. The intensity-adaptive weight \(w_{p,q} = \exp(-|I(p) - I(q)|/\sigma_w)\) controls the strength of the pairwise term depending on the intensity difference between two neighboring pixels.

The co-planar potential \(\phi_{\text{co}}\) calculates the average difference of the homography motions for all pixels within the union of the two superpixels, as they are on the same plane by assumption. The hinge potential \(\phi_{h}\) penalizes the motion difference only for the middle pixel and applies a constant bias \(\lambda_{h}\). For handling the remaining cases such as occlusion or disocclusion, the truncation constant \(\tau_{p}\) is used. Note that the pairwise cost only becomes effective between neighboring pixels that belong to two different superpixels. The cost between neighboring pixels in the same superpixel is naturally zero, because the two pixels have the same associated homography motion.

The pairwise occlusion term simply encourages the spatial smoothness of the occlusion states by penalizing their differences between neighboring pixels. If neighboring pixels have differing occlusion states, the term incurs the constant penalty \(\lambda_{o}\).

### 3.2.3 Forward-backward consistency term

Unlike conventional optical flow algorithms, our model explicitly integrates forward-backward consistency into the objective function \([20]\):

\[
E_{\text{fc}}(H^f, H^b, o^f, o^{f+1}) = \sum_{p \in I^f} C^f_p + \sum_{p \in I^{f+1}} C^{f+1}_p \tag{5a}
\]

with

\[
C^f_p = \delta_p o^f \cdot \rho_c(\|p - H^b_{\text{sp}} p_{\text{sp}}\|) \tag{5b}
\]

\[
C^{f+1}_p = \delta_p o^{f+1} \cdot \rho_c(\|p - H^f_{\text{sp}} p_{\text{sp}}\|) \tag{5c}
\]

where \(p' = H^b_{\text{sp}} p\) and \(p'' = H^b_{\text{sp}} p\). The consistency term penalizes the Euclidean distance between the position of pixel \(p\) and the back-projected position of the corresponding pixel in the other frame, as illustrated in the first column of Fig. 3. Thus, the term encourages the homography matrices from the two corresponding points to be under an inverse relationship, providing a soft constraint on motions in the opposite view, which improves the estimation as shown below. The function \(\rho_c(\cdot)\) truncates its input at \(\tau_{C}\) to be robust to possible outliers of flow or occlusion.

The term takes into account all pixels in both views, but applies the penalty only if the pixel \(p\) and its corresponding pixel \(p'\) or \(p''\) in the respective other frame are not occluded. This condition enforces pixels to be either occluded or their motion to satisfy the bi-directional symmetry property. The condition may lead to a trivial solution in which all pixels are marked as becoming occluded such that no penalties arise from this term. However, the constant penalty \(\lambda_{\text{ooc}}\) in the data term, c.f. Eq. (2b) and Eq. (2c), prevents the solution from falling into this trivial case and balances between visible pixels and occlusions.

### 3.2.4 Occlusion-disocclusion symmetry term

The occlusion-disocclusion symmetry term plays the most important role in our model, and allows flow and occlusions to mutually leverage one another. Specifically, the term penalizes cases in which the occlusion-disocclusion symmetry relationship does not hold in both views:

\[
E_{\text{oc}}(o^f, o^{f+1}) = \sum_{p \in I^f} d_p o^f \cap N_p^f + \sum_{p \in I^{f+1}} d_{p^{f+1}} o^{f+1} \cap N_{p^{f+1}} \tag{6a}
\]

with

\[
N_p^f = \left\{ p \mid p = H_{\text{sp}}^f p', \forall p' \in I^{f+1} \right\} \tag{6b}
\]

\[
N_{p^{f+1}} = \left\{ p \mid p = H_{\text{sp}}^b p', \forall p' \in I^f \right\}. \tag{6c}
\]

The XNOR operation (e.g., \(0 \odot 0 = 1\) and \(1 \odot 0 = 0\)) ensures that if a pixel \(p\) is being occluded in one frame, then there cannot be any pixels in the other frame whose motion map to \(p\). We explicitly detect disocclusion in each view, representing it with the variables \(N_p^{f+1}\) and \(N_{p^{f+1}}\), respectively. \(N_p^{f+1}\) denotes the number of pixels that are mapped to \(p\) in \(I^f\) when warping pixels in \(I^{f+1}\) to \(I^t\) given their corresponding motion \(H_{sp}^f\). \(N_{p^{f+1}}^{f+1}\) is defined analogously by warping pixels from \(I^t\) to \(I^{f+1}\) given the set of per-superpixel homography motions \(H_{sp}^b\). If no pixel is mapped to \(p\) in \(I^{f+1}\) (i.e., \(N_p^{f+1} = 0\)), the pixel \(p\) is being disoccluded, as visualized in the third column of Fig. 3. By defining disocclusions solely as a result of the motion, their corresponding variables do not need to be optimized directly, but are indirectly determined by the motion \(H_{sp}^b\) and \(H_{sp}^f\).

In previous studies \([22, 41]\), similar concepts have been introduced in different forms by encouraging unique correspondences between pixels visible in both views. Instead,
3.3. Optimization

We jointly optimize the two different sets of variables – homography motions (i.e., $H^f$ and $H^b$) and occlusion maps (i.e., $o^f$ and $o^{f+1}$) – using a block coordinate descent algorithm, which optimizes the variables alternatingly. As described in Algorithm 1, we first estimate the forward flow $H^f$. Then, we update $N^{t+1}$ from forward flow $H^f$ and estimate the occlusion map $o^{f+1}$. The remaining variables $H^b$, $N^t$, and $o^t$ are updated in turn in a similar manner.

3.3.1 Optimizing $H^f$ and $H^b$

Due to the difficulty of optimizing the continuous variables (e.g., stemming from the nonlinearity of the data term), we instead solve a discrete multi-label optimization problem that collects a number of candidate homography motions as proposal sets and then chooses the most suitable motion for each superpixel using fusion moves or $\alpha$-expansion with QPBO [24, 32, 39, 44]. For an efficient optimization, we sequentially run expansion moves locally on subgraphs of superpixels (e.g., a set of neighboring 30 superpixels with 70% overlap between each other)\(^1\) and then globally on all the superpixels as described in Algorithm 2. This strategy follows the similar approach in [40], which defines subgraphs on multiple scales and optimizes them sequentially. We found that this combined local and global optimization strategy yields a faster convergence while avoiding local minima, especially for the planar surface representation with homography-parameterized motions as used here.

3.3.2 Optimizing $o^t$ and $o^{t+1}$

Optimizing the occlusion maps is relatively simple. For instance, once $N^t$ is updated from the backward flow $H^b$, the binary occlusion map $o^t$ can be estimated via graph-cuts [6, 7, 23] while holding all other variables fixed. The pairwise occlusion term in Eqs. (3b) and (3c) is submodular, thus making standard graph-cuts applicable.

4. Experiments

We evaluate the accuracy of our algorithm on the KITTI Optical Flow 2015 benchmark [12] and on the MPI Sintel Flow Dataset [8], both qualitatively and quantitatively. Additionally, we analyze the importance of our symmetric

---

\(^1\)See supplementary material for an analysis of this design choice.
approach by turning each term off and evaluating how significantly it affects the accuracy. For faster convergence, we use DiscreteFlow [29] to initialize our estimation as well as to derive proposals. We automatically tune the parameters (weights $\lambda$, truncation thresholds $T_D$, $T_p$, $T_C$, and biases $\lambda_{occ}$, $\lambda_{bg}$) using Bayesian optimization [27] on the training portion of each benchmark.

4.1. KITTI Optical Flow 2015

On the KITTI Optical Flow 2015 benchmark, our MirrorFlow algorithm outperforms all two-frame optical flow methods at the time of writing, demonstrating the lowest percentage of flow outliers ($Fl\text{-all}$). Table 1 gives detailed numbers, Fig. 4 shows qualitative results. Leveraging our symmetry terms and motion representation based on piecewise homographies, our method also demonstrates the second best results for handling flow on dynamic foreground objects ($Fl\text{-fg}$) and handling background motion ($Fl\text{-bg}$).

For evaluating the accuracy especially in occluded areas, we cannot directly rely on the KITTI accuracy indicators, as the number of occluded pixels and non-occluded pixels in the testing dataset are unknown. However, by looking at the gap of outlier percentages between all pixels and non-occluded pixels, we can infer that our symmetric method is among the most accurate (probably the most accurate) method in occluded areas (see supplementary material for a more detailed discussion).

One important remark is that our method does not use any learned feature descriptors or semantic information unlike other top-performing algorithms [2, 19, 33, 46, 48]. Even without them, our method significantly outperforms the baseline method [29], which is the next best method not relying on learned descriptors or semantics, and is also used for initialization. We significantly reduce the number of incorrect pixels by more than 50% by virtue of our joint, symmetric formulation. We believe that our method still has room for substantial further improvement by exploiting semantic information or using learned feature descriptors; we leave this for future work.

<table>
<thead>
<tr>
<th>Method</th>
<th>Non-occluded pixels</th>
<th>All pixels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Fl\text{-bg}$</td>
<td>$Fl\text{-fg}$</td>
</tr>
<tr>
<td>Ours (MirrorFlow)</td>
<td>6.24 %</td>
<td>12.95 %</td>
</tr>
<tr>
<td>FlowNet2 [19]</td>
<td>7.22 %</td>
<td>5.60 %</td>
</tr>
<tr>
<td>SDF [3]</td>
<td>5.75 %</td>
<td>18.38 %</td>
</tr>
<tr>
<td>MR-Flow [46]</td>
<td>6.86 %</td>
<td>17.91 %</td>
</tr>
<tr>
<td>DCFLOW [48]</td>
<td>8.04 %</td>
<td>19.84 %</td>
</tr>
<tr>
<td>SOF [33]</td>
<td>8.11 %</td>
<td>18.16 %</td>
</tr>
<tr>
<td>DiscreteFlow [29]</td>
<td>9.96 %</td>
<td>17.03 %</td>
</tr>
</tbody>
</table>

Table 1. KITTI Optical Flow 2015: Comparison to top-performing optical flow algorithms in the benchmark in terms of percentages of pixels with an incorrect flow estimate (at the default threshold of 3 pixels). The best and the second best results are marked in bold and underlined, respectively.

4.2. MPI Sintel Flow Dataset

While we focus on the KITTI dataset with its challenging scenes in the context of autonomous driving, we also evaluate our approach on the MPI Sintel Flow Dataset [8], where approaches based on piecewise rigidity are known to be somewhat disadvantaged. Nevertheless, our method still performs rather competitively, achieving the second place in the Clean pass and the 13th place in the Final pass, both at the time of writing. Table 2 gives detailed accuracy numbers. Fig. 5 shows qualitative results using the color code of the Sintel dataset, where we observe rather accurate estimates of occluded regions. The main reason why our method is not as accurate as on the KITTI benchmark is that our planar-rigid motion assumption is not as appropriate for the Sintel Dataset, where the majority of motions are non-rigid and most surfaces are non-planar. Despite of this limitation, our method demonstrates leading results especially on occluded pixels in the Clean pass, which once again confirms the key benefits of our joint flow and occlusion estimation pipeline. Note that while we do not consider this here, the key ideas behind our joint, symmetric approach are not limited to piecewise rigid representations.

<table>
<thead>
<tr>
<th>Method</th>
<th>Final pass</th>
<th>Clean pass</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EPE</td>
<td>EPE</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>noccl.</td>
</tr>
<tr>
<td>DCFLOW [48]</td>
<td>5.119</td>
<td>2.283</td>
</tr>
<tr>
<td>MR-Flow [46]</td>
<td>5.376</td>
<td>2.818</td>
</tr>
<tr>
<td>S2F-IF [52]</td>
<td>5.417</td>
<td>2.549</td>
</tr>
<tr>
<td>Ours (MirrorFlow)</td>
<td>6.071</td>
<td>3.186</td>
</tr>
</tbody>
</table>

Table 2. MPI Sintel Flow Dataset: Accuracy in terms of the average end-point error (EPE). Leading algorithms on Final or Clean. Our model performs the second best in the Clean pass.

4.3. Importance of symmetries

We conduct an ablation study to emphasize the contribution of our symmetric formulation and to demonstrate how much each symmetry property contributes to the flow estimation accuracy. As a baseline, we first consider an asymmetric version of our model, which only relies on the data term and the pairwise term (Asymm). Then, we extend it to the symmetric case by estimating flow bi-directionally through enabling the forward-backward consistency term (Symm+c) and the occlusion-disocclusion symmetry term (Symm+o). The full model has both these terms enabled (Symm+c+o). We conduct this study on the KITTI Optical Flow 2015 training set.

As shown in Table 3, the occlusion-disocclusion symmetry term has the most significant contribution to the accuracy of flow estimation. The error decreases substantially
Figure 4. **Qualitative results on KITTI 2015**: (left to right) ground truth flow, forward flow, backward flow, occlusion map overlayed on the current frame and the next frame, respectively. Note that the ground truth flow map on KITTI is sparse, and some objects are masked.

Figure 5. **Qualitative results on Sintel**: (left to right) ground truth flow, forward flow, backward flow, ground truth occlusion, occlusion maps, overlayed on the current frame and the next frame, respectively. Note the high agreement between true and estimated occlusions.

<table>
<thead>
<tr>
<th>Method</th>
<th>Non-occluded pixels</th>
<th>All pixels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fl-bg</td>
<td>Fl-fg</td>
</tr>
<tr>
<td>Symm+cs</td>
<td>6.52 %</td>
<td>11.72 %</td>
</tr>
<tr>
<td>Symm+s</td>
<td>6.73 %</td>
<td>11.90 %</td>
</tr>
<tr>
<td>Symm+c</td>
<td>10.41 %</td>
<td>18.17 %</td>
</tr>
<tr>
<td>Asymm</td>
<td>8.39 %</td>
<td>14.97 %</td>
</tr>
</tbody>
</table>

Table 3. **Ablation study for each term** on KITTI 2015 training: The forward-backward consistency term (c) and the occlusion-disocclusion symmetry term (s) both play important roles in our symmetric model. Their combination (cs) shows the best results.

by about 20%. This observation follows one of our motivations that the accurate localization of occluded areas contributes to estimating flow more accurately. The forward-backward consistency term boosts the quality of the results further, but only when the symmetry term is turned on. This is because the forward-backward consistency relies on accurate estimates of the occlusion regions, which are only available when the occlusion-disocclusion symmetry is considered as well. When both terms are active we achieve the best accuracy, which highlights the benefit of our full symmetric pipeline.

5. Conclusion & Future Work

We have proposed a symmetric optical flow method that jointly estimates optical flow in both directions and occlusion maps for each view by exploiting the symmetry properties that they possess in the two consecutive images. We exploit both forward-backward consistency of the flow as well as occlusion-disocclusion symmetry, and formulate a piecewise rigid model. Our results on widely-used public optical flow benchmarks clearly demonstrate that our joint, symmetric approach yields significant improvements in flow estimation accuracy, especially in occluded areas. For the challenging KITTI benchmark, we report leading results even without employing any semantic knowledge or learning of appearance descriptors. We believe that a superpixel refinement, employing non-rigidity on top of our rigid motion model, or utilizing learned appearance descriptors will lead to further improvements in the future.

**Acknowledgement.** The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007–2013) / ERC Grant Agreement No. 307942.
References


