Dense, Accurate Optical Flow Estimation with Piecewise Parametric Model

Jiaolong Yang [1,2] and Hongdong Li [2,3]

[1]Beijing Lab of Intelligent Information Technology, Beijing Institute of Technology
[2]Research School of Engineering, The Australian National University (ANU) and NICTA
[3]ARC Centre of Excellence for Robotic Vision (ACRV)

Abstract

This paper proposes a simple method for estimating dense and accurate optical flow field. It revitalizes an early idea of piecewise parametric flow model. A key innovation is that, we fit a flow field piecewise to a variety of parametric models, where the domain of each piece (i.e., each piece's shape, position and size) is determined adaptively, while at the same time maintaining a global inter-piece flow continuity constraint. We achieve this by a multi-model fitting scheme via energy minimization. Our energy takes into account both the piecewise constant model assumption and the flow field continuity constraint, enabling the proposed method to effectively handle both homogeneous motions and complex motions. The experiments on three public optical flow benchmarks (KITTI, MPI Sintel, and Middlebury) show the superiority of our method compared with the state of the art: it achieves top-tier performances on all the three benchmarks.

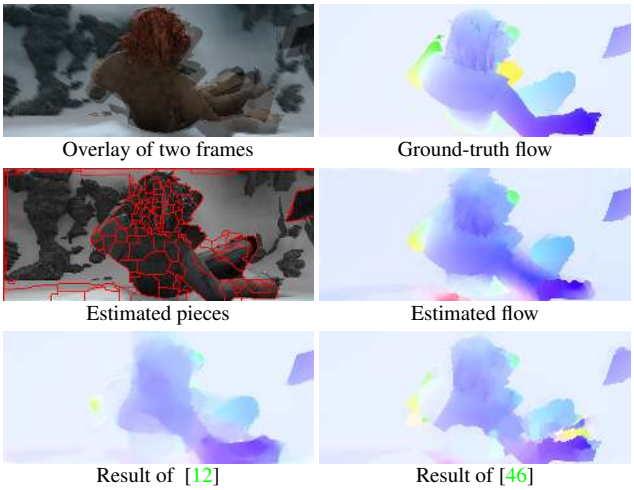


Figure 1: The proposed method estimates optical flow using piecewise parametric (homography) models. In this example it yields accurate motion estimate on the actor's shoulder and back compared to LDOF [12] and MDP-OF [46] .

1. Introduction

As a classic topic in computer vision, optical flow computation has attracted considerable attentions from the community. Remarkable progresses have been made in the past decades, with high performance optical flow algorithms available nowadays [11, 48, 37, 46, 25]. Despite these successes, to obtain dense and accurate flow field remains challenging, especially for general dynamic scenes containing multiple complex, non-rigid objects,and/or large motions.

This paper revisits the idea of piecewise parametric optical flow estimation popularized by Black *etc*. researchers in the 1990s [7, 6, 23]. Unlike most modern optical flow techniques which capitalize on dense per-pixel flow vector estimation, these piecewise parametric flow methods assume a low-order parametric motion models within each segmented image piece. Using parametric models to represent a flow field, while is compact, can be rather restrictive. When the motion field is very complex, or when image segments do not conform well to motion segments, the parametrically-fitted flow field can be inaccurate or erroneous. Partly due to this reason, piecewise parametric models are seldom adopted by modern optical flow methods [37].

In this paper, we advocate that, equipped with a carefully devised energy function and modern minimization techniques, the piecewise parametric model can be revitalized to achieve highly-accurate optical flow estimation with state-of-the-art performance.

Our motivation is described as follows. As in previous work [7, 23], we assume that a flow field can be jointly represented by multiple parametric motion models in a piecewise fashion. To ease description, the 8-dof homography transformation model is used in this paper. To achieve accurate model fitting or approximation, we allow the size

and shape of each piece to change adaptively. For example, some pieces must be large to account for large regions with homogeneous motion vectors to gain fitting robustness, while others need to be small enough to capture fine motion details within a small region containing complex motions. The approach of this work is to determine each piece appropriately, and at the same time to fit a parametric model to each piece (see Fig. 1 for illustration). In light of this, the proposed method is similar to the joint motion estimation and motion segmentation scheme, as investigated in *e.g.* [15, 5, 38, 39, 40].

However, there is a subtle but critical difference. In contrast to the above methods which aimed to segment a motion field into a few independently moving regions [15, 5, 38], our aim is to fit the entire flow field with a large number of (possibly hundreds of) piecewise parametric models. The proposed method can effectively handle complex motions which are challenging for the above methods such as [5, 40].

The contributions of this work are threefold. ***i)*** We represent and estimate the flow field with piecewise homography models, and solve the problem via joint discrete and continuous optimization similar to the multi-model fitting work of Isack and Boykov [22]. ***ii)*** We propose a novel energy formulation where the regularization takes into account of both a piecewise constant model constraint and a flow field continuity constraint. It allows the method to yield large pieces for homogeneous motion, and small pieces for complex motion. ***iii)*** We show that the piecewise parametric model can be modified to estimate flow field accurately. At the time of writing, our method achieves the top accuracy on the KITTI benchmark [20], outperforms all published methods on the MPI Sintel (clean) benchmark [13], and yields top-tier results on the Middlebury benchmark [1].

1.1. Related work

There is a large volume of work on optical flow estimation. Below we only review related methods for piecewise segmentation based and/or parametric flow estimation.

Computing parametric flow field on pre-segmented images is a classic idea [7, 23, 45, 27]. Black and Jepson [7] segment the image with color cue and fit variable-order motion model to each segment independently. Ju *et al.* [23] divide the image evenly into rectangular patches, and fit affine models to them. Interaction between neighbouring patches is involved and defined to be the difference of model parameters. Xu *et al.* [45] fit affine models on regions segmented with color cue and initial flow; the fitting is regularized with Total Variation of the flow field. Lei and Yang [27] represent the image with region-tree built from color segmentation; constant flow vector is computed for each region.

Another category of related methods first estimates candidate parametric models then assign these models to each pixel as a segmentation process, *e.g.* [44, 4, 14, 41]. Wills *et al.* [44] use multiple homographies fitted from feature matches for segmentation, and Bhat *et al.* [4] use both homographies and fundamental matrices. Recently, Chen *et al.* [14] use translation and similarity transformations extracted from nearest neighbour field for segmentation. In scene flow estimation, Vogel *et al.* [41] assign each pixel a segment, and each segment a 3D plane; the plane candidates are fitted based on an input scene flow estimate.

Methods have been proposed for joint motion segmentation and estimation [29, 15, 35, 50, 5, 47, 40]. For example, Mémin *et al.* [29] proposed such an approach in a variational framework, however the energy is defined on incremental motion field during the coarse-to-fine processing. Cremers and Soatto [15] developed a variational approach to jointly estimate segmentation boundaries and affine models via continuous optimization. Roussos *et al.* [35] represent and estimate dense motion field via multiple fundamental matrices plus an inverse-depth field.

Layered model estimation is another useful technique for motion segmentation and estimation [42, 38, 39]. This approach estimates a few overlapping motion layers, typically represented by parametric models, and assigns pixels to these layers. The pioneer work of Wang and Adelson [42] uses affine layers to represent the motion field, and recent advances by Sun *et al.* [38, 39] use affine motion to regularize the flow in each layer. The motivation of these approaches and their formulations are different from ours.

The proposed method is related to the methods based on over-parameterization [31, 21]. Nir *et al.* [31] represent optical flow with parametric (*e.g.* affine and rigid) model defined on every pixel. A recent work by Hornáček *et al.* [21] defines per-pixel homography for flow estimation. In contrast to pointwise parametric model, our method fits piecewise constant parametric models on adaptive segments.

The optimization scheme we use is similar to the multi-model fitting work [22], and other relevant works in different domains, *e.g.* [36, 32]. Compared to [22] where scattered data is fitted to each model independently, we deal with dense, regular image grids where the models interact with each other to address the spatial continuity of flow field. Our idea of adaptively changing the domains of image pieces is partly related to the works of image quilting [19] and photo autocollage [34].

2. Piecewise Parametric Flow Estimation

Given two images frames $\mathbf{I}_1$ and $\mathbf{I}_2$ as the reference frame and target frame respectively, our goal is to estimate a dense 2D displacement vector $\mathbf{u}$ at each pixel $\mathbf{x}$ on $\mathbf{I}_1$, based on the the brightness constancy assumption, *i.e.* $\mathbf{I}_1(\mathbf{x}) = \mathbf{I}_2(\mathbf{y})$ where $\mathbf{y} = \mathbf{x} + \mathbf{u}$. The displacement vector can be represented by a parametric transformation model T, *i.e.*, $\mathbf{y} = \mathrm{T} \circ \mathbf{x}$. In this paper, we choose to use the 8-dof ho-

mography as the parametric model, although other types of parametric models are also possible. One obvious benefit of choosing the homography model is that, homographies can be induced by 3D planes undergoing rigid motion. In fact, even for certain non-rigid motions or deformations, homography can be used as a good transformation model.

2.1. Energy function

Let $\mathcal{L} = \{1, .., K\}$ be a set of discrete labels representing the set of K homography models, *i.e.* $\mathcal{H} = \{\mathrm{H}_k\}, k = 1, \ldots, K$. Let Ω be the 2D image domain of $\mathbf{I}_1$, and $L : \Omega \rightarrow \mathcal{L}$ be a labeling function. Assigning label $k = L(\mathbf{x})$ to pixel $\mathbf{x}$ means that motion of $\mathbf{x}$ is induced by homography $\mathrm{H}_k \in \mathcal{H}$.

Our energy function is defined on both the unknown piecewise parametric models $\mathcal{H}$, and the unknown pixel labelling L, as

$$E(\mathcal{H}, L) = E_D(\mathcal{H}, L) + \lambda_C E_C(\mathcal{H}, L) + \lambda_P E_P(L) + \lambda_M E_M(L), \quad (1)$$

where E_D is a data term, E_C is a flow continuity regularization term, E_P is a Potts model term, and E_M is a *label cost* term [28, 16] reflecting the Minimum Descriptor Length (MDL) principle. The λs are weighting parameters. Note that, one homography model can be be assigned to multiple disjoint pieces, as this is beneficial to handel occasion.

2.2. Data term

The data term E_D enforces the brightness constancy constraint, subjecting to the piecewise homography models as

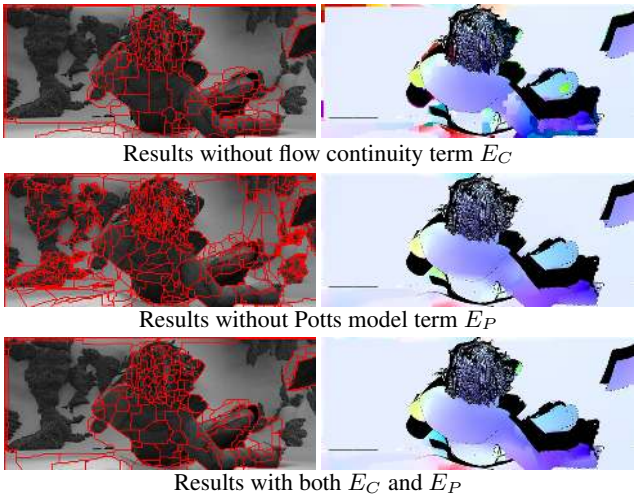
$$E_D(\mathcal{H}, L) = \sum_{\mathbf{x} \in \Omega} |\mathbf{I}_1(\mathbf{x}) - \mathbf{I}_2(\mathrm{H}_{L(\mathbf{x})}\mathbf{x})|, \quad (2)$$

where $|\cdot|$ denotes the L_1 norm. For brevity, we slightly abuse notations hereafter: H needs to be understood as an operator rather than matrix; both $\mathbf{x}$ and H$\mathbf{x}$ represent inhomogeneous image coordinates.

To improve the robustness with respect to noise and illumination changes, we use a robustified data term as in [11, 8]. The robust version takes into account of both brightness constancy constraint and gradient constancy constraint, in addition to the use of a robust estimator ρ_D:

$$E_D(\mathcal{H}, L) = \sum_{\mathbf{x} \in \Omega} \rho_D\big((1-\alpha)|\mathbf{I}_1(\mathbf{x}) - \mathbf{I}_2(\mathrm{H}_{L(\mathbf{x})}\mathbf{x})| + \alpha|\nabla\mathbf{I}_1(\mathbf{x}) - \nabla\mathbf{I}_2(\mathrm{H}_{L(\mathbf{x})}\mathbf{x})|\big), \quad (3)$$

where we choose ρ_D to be a truncating function as $\rho_D(\cdot) = \min(\cdot, \sigma_D)$ and σ_D is a scalar parameter.



Results without flow continuity term E_C

Results without Potts model term E_P

Results with both E_C and E_P

Figure 2: Effects of E_C and E_P. **Top row:** without E_C, the estimated flow contains many gross errors on the foreground human body with complex motion. (Occluded regions are masked black) **Middle row:** without E_P, the background regions with homogeneous motion are not well grouped, leading to 0.05∼0.1 endpoint error increase for them. **Bottom row:** with both the two terms, the method handles well both complex and homogeneous motions.

2.3. Flow continuity (inter-piece compatibility) term

We introduce a flow continuity term E_C, which enforces the continuity constraint of the flow field, rather than the widely-used 1st-order or 2nd-order smoothness constraint (*e.g.* TV [48] or TGV [10] regularizer). Let $\mathcal{E}$ be the set of 4-connected pixel pairs on the image, E_C is defined to be

$$E_C(\mathcal{H}, L) = \sum_{(\mathbf{x}, \mathbf{x}') \in \mathcal{E}} w(\mathbf{x}, \mathbf{x}') \cdot \rho_C\big(|\mathrm{H}_{L(\mathbf{x})}\bar{\mathbf{x}} - \mathrm{H}_{L(\mathbf{x}')}\bar{\mathbf{x}}|\big), \quad (4)$$

where $\bar{\mathbf{x}} = (\mathbf{x} + \mathbf{x}')/2$ is the midpoint of $(\mathbf{x}, \mathbf{x}')$, $\rho_C(\cdot) = \min(\cdot, \sigma_C)$ with σ_C a scalar parameter, and $w(\mathbf{x}, \mathbf{x}') = \exp(-\beta\|\mathbf{I}_1(\mathbf{x}) - \mathbf{I}_1(\mathbf{x}')\|)$ is a color-based weighting term. Note that if $L(\mathbf{x}) = L(\mathbf{x}')$, the cost at pixel-pair $(\mathbf{x}, \mathbf{x}')$ is nil. The properties of this term are analyzed as follows.

- E_C does not penalize the variations between neighbouring pixels within a single piece (where all interior pixels have the same label), even if the variations are large. It only penalizes motion discontinuities at inter-piece boundaries (hence we also call it the *inter-piece compatibility term*).

- The inter-piece motion discrepancies can be 0 or very small (*i.e.* the two adjacent pieces are compatible) even if their homography models differ a lot. Thus E_C allows model-switch, which is important for handling complex motion.

- It is easy to see that E_C is a sub-modular function in terms of discrete labeling variables L, which is a nice property for discrete energy minimization [26].

The effect of this term is illustrated in Fig. 2. It can be seen that without E_C the estimated flow contains many sharp discontinuous and gross errors.

2.4. Potts model term

In addition to the pairwise flow continuity term E_C, we use a pairwise Potts model term E_P to encourage spatially coherent labeling. This term is defined only on the discrete labeling variables as $E_P(L) = \sum_{(\mathbf{x},\mathbf{x}')\in\mathcal{E}} \delta\big(L(\mathbf{x}) \neq L(\mathbf{x}')\big)$, where $\delta(\cdot)$ is the 0-1 indicator function which takes 1 if the input argument is true, and 0 otherwise.

The terms E_C and E_P have different effects, and are complementary to each other. E_P enforces *intra-piece model constancy*; it penalizes any model change, no matter how similar the two models are. In contrast, as mentioned before, E_C enforces *inter-piece motion compatibility*; it allows compatible model switch, no matter how different the two models are (cf. Sec. 2.3).

Figure 2 illustrates that, without E_P the regions with homogeneous motion are not well grouped. This may lead to inferior flow estimate for these regions. Moreover, this can also be harmful to other regions: a model can be accidently assigned to many small pieces during labeling, bringing in difficulties for model estimation (cf. Sec. 3.1).

2.5. MDL term

To reduce the redundancy of the fitted homographies, we employ an MDL term E_M to penalize the total number of the used homography models, *i.e.* $E_M(L) = \sum_{k=1}^{K} \tau(k)$, where $\tau(k) = \begin{cases} 1, & \text{if } \sum_{\mathbf{x}\in\Omega} \delta(L(\mathbf{x}) = k) > 0 \\ 0, & \text{othewise} \end{cases}$.

This term is helpful especially when a prior knowledge exists that the flow field can be well approximated by a relatively small number of parametric models. For example, in some man-made scenes where there are large planar structures, this term helps encourage less homographies and increase fitting quality. One may set its weight λ_M to 0 or very small if no prior is given.

3. Optimization

This section presents our optimization techniques. We first present the alternation based optimization assuming initial parameters given, then show our initialization method.

3.1. Alternation

The energy defined in Eq. 1 involves both discrete variables L and continuous variables $\mathcal{H}$. We approach this discrete-continues problem similarly to the multi-model fitting method of [22]. A block coordinate descent (see Algorithm 1) is used that alternates between optimizing over L and $\mathcal{H}$, thus splitting the original problem into two subproblems described as follows.

Algorithm 1: Piecewise Homography Flow

```
Initialize H, L.
while not converge do
    Fix H, solve for L in Eq. 5 via graph-cut [17].
    Fix L, solve for H via Algorithm 2.
```

Algorithm 2: Piecewise Homography Fitting

```
Sort the input homographies H_k, k=1,...,K according
to their labeling area in L in descending order.
for iteration = 1, ..., m do
    for k = 1 : K do
        Optimize H_k in Eq. 6 by simplex downhill [30].
```

I. Labeling: Solve for L with fixed $\mathcal{H}$. With fixed homographies, the energy minimization reduces to a labeling problem with the energy

$$E(L) = \underbrace{E_D(L)}_{\substack{\text{Unary}\\ \text{potential}}} + \underbrace{\lambda_C E_C(L) + \lambda_P E_P(L)}_{\substack{\text{(Submodular) Pairwise}\\ \text{potential}}} + \underbrace{\lambda_M E_M(L)}_{\substack{\text{MDL}\\ \text{potential}}}. \quad (5)$$

Without the MDL term, the energy corresponds to a standard Markov Random Field (MRF) problem with unary and pairwise potentials. The α-expansion based graph-cut method [9] can be used for fast approximate energy minimization. We use the method of [17] to handle the label costs in the MDL term.

A large set of homography models (*e.g.*, 1,000 in our experiments) are generated during initialization (See Sec. 3.2). For the sake of computational efficiency, if a homography is not labeled to any pixel after one round of the labeling process of L, it will be removed from the candidate model set. Another strategy to speed up computation is restricting the α-expansion within a region of limited radius on the image plane (*e.g.*, $<$100 pixels).

II. Fitting: Solve for $\mathcal{H}$ with fixed L. The homography parameters $\mathcal{H}$ appear in the data term E_D and flow continuity term E_C. With fixed labeling, minimizing the energy function is an unconstrained continuous optimization problem. If $\mathcal{H}$ appears only in E_D, we can optimize the parameters of each homography independently. Unfortunately, it appears also in E_C which involves pairwise iterations between adjacent pieces. To tackle this issue we propose to

use an inner block coordinate decent procedure: the homography is optimized one by one, each time with other homographies fixed. The homography models with larger labelling areas are first optimized as they are generally less affected by E_C. See Algorithm 2.

When optimizing a homography H_k, the energy reads

$$\begin{aligned} E(\mathrm{H}_k) &= E_D(\mathrm{H}_k) + \lambda_C E_C(\mathrm{H}_k) \\ &= \sum_{\mathbf{x}\in\Omega_k} \rho_D\big((1-\alpha)|\mathbf{I}_1(\mathbf{x})-\mathbf{I}_2(\mathrm{H}_k\mathbf{x})|+\alpha|\nabla\mathbf{I}_1(\mathbf{x})-\nabla\mathbf{I}_2(\mathrm{H}_k\mathbf{x})|\big) \\ &\quad +\lambda_C \sum_{(\mathbf{x},\mathbf{x}')\in\mathcal{E}_k} w(\mathbf{x},\mathbf{x}')\cdot\rho_C\big(|\mathrm{H}_k\bar{\mathbf{x}} - \mathrm{H}_{L(\mathbf{x}')}\bar{\mathbf{x}}|\big) \end{aligned} \tag{6}$$

where $\Omega_k = \{\mathbf{x}\in\Omega \mid L(\mathbf{x}) = k\}$, $\mathcal{E}_k = \{(\mathbf{x},\mathbf{x}')\in\mathcal{E} \mid L(\mathbf{x}) = k,\ L(\mathbf{x}') \neq k\}$, and other variants and functions are as in Eq. 3 and Eq. 4. We optimize Eq. 6 via the derivative-free Nelder-Mead Simplex Downhill method [30]. Similar to Zhang *et al*. [49], the vertexes of a simplex are initialized with the homographies of adjacent pieces. We found this strategy to be very effective in reducing the energy.

3.2. Initialization

To generate candidate homography proposals and an initial labelling, we use PatchMatch [3] to compute an initial correspondence field. Then we use DLT to fit homographies for small (*e.g.* 5×5) local regions, and grow the regions to consistent neighbouring pixels for initial labelling. See *Supplementary Material* for more details.

4. Post-processing

4.1. Occlusion handling

We detect occlusion based on the forward-backward consistency check. A pixel $\mathbf{x}$ will be considered as occluded if $\|\mathbf{x} - \mathrm{H}'_{l'}\mathrm{H}_l\mathbf{x}\| > \theta$, where $\mathrm{H}'_{l'}$ is the homography of the point $\mathrm{H}_l\mathbf{x}$ on the target image, and θ is a scalar threshold. We then remove the data term E_D in Eq. 1, and label estimated homographies to the occluded pixels via graph-cut.

4.2. Refinement

To further improve the results for complex motion, small local deformation may be necessary to compensate the discrepancy between true flow field and the piecewise approximation. Therefore we use the publicly available code of the "Classic+NL-fast" method [37] for flow refinement[1]. Note that we directly refine the flow on the original image scale, and no coarse-to-fine pyramid structure is used.

5. Evaluation

In this section, we test the proposed method on three public benchmarks: the KITTI flow benchmark [20], the MPI Sintel benchmark [13], and the Middlebury flow benchmark [1]. Our method is implemented in C++ & Matlab, and tested on a standard PC with Intel i7 3.4GHz CPU. In the following experiments, we set $\alpha = 0.9$, $\beta = 5$, $\sigma_D = 10$, iteration number of Algorithm 1 to be 5, maximal iteration of Algorithm 2 to be 15. Other parameters are trained on the benchmarks and will be explained in the corresponding sections. During initialization we allow a maximum number of $1,000$ pieces. The images on KITTI and Sintel are half-sized before running the method.

Table 1: End-point Error results on part of the *training* sequences in KITTI benchmark (3-pixel error threshold).

	Out-Noc	Out-All	Avg-Noc	Avg-All
Without MDL	5.94 %	11.44 %	1.58	3.69
Without refinement	5.76 %	10.84 %	1.41	3.00
Full	5.56 %	10.81 %	1.36	2.98

5.1. Results on KITTI

The KITTI dataset is a challenging real-world dataset containing non-lambertian surfaces, different lighting conditions and large displacement motions.

We selected 20 image pairs with ground-truth flow fields from the *training* set. After training we set $\lambda_C = 1$, $\lambda_P = 4$, $\lambda_M = 400$, $\sigma_C = 10$, and $\theta = 1.5$. The results on these images are shown in Tab. 1, where "Out-Noc" and "Avg-Noc" refer respectively to the outlier ratio and average end-point error in non-occluded regions and "Out-All" and "Avg-All" to all regions with ground-truth. The effect of the MDL term is obvious on this benchmark. Table 1 shows that the MDL term improves the results obtained without MDL term (*i.e.* $\lambda_M = 0$) by about $10\% \sim 20\%$. Figure 3 presents the estimated pieces with different MDL weights. We found that $40 \sim 80$ homography models are generally adequate for flow estimation on this dataset. Table 1 also shows that the refinement step improves the results by around $3\% \sim 5\%$.

We ran our method on the *test* set where the ground-truth is hidden. Figure 4 shows two examples of our homography motion segmentation and flow estimation results. Note that both the large surfaces of road, green belt, building facades, cars, *etc*., and the small objects such as road lamp and sign are well segmented. Table 2 compares our results with state-of-the-art two-frame optical flow methods. At the time of writing, our method is ranked the first among all published methods, under the by default 3-pixel threshold metric. In fact, our method shows improved performance on almost all the reported metrics used in KITTI.

5.2. Results on Middlebury

The Middlebury optical flow benchmark only contains relatively small displacements. It has been extensively studied in recent years and sub-pixel accuracy has been

[1]The refinement using this method yields worse results in occluded regions; we keep the original flow for pixels that are very likely occluded (which failed the forward-backward check with a large threshold $\theta = 20$).

$\lambda_M = 0$ (135 models) $\lambda_M = 300$ (56 models) $\lambda_M = 600$ (36 models)

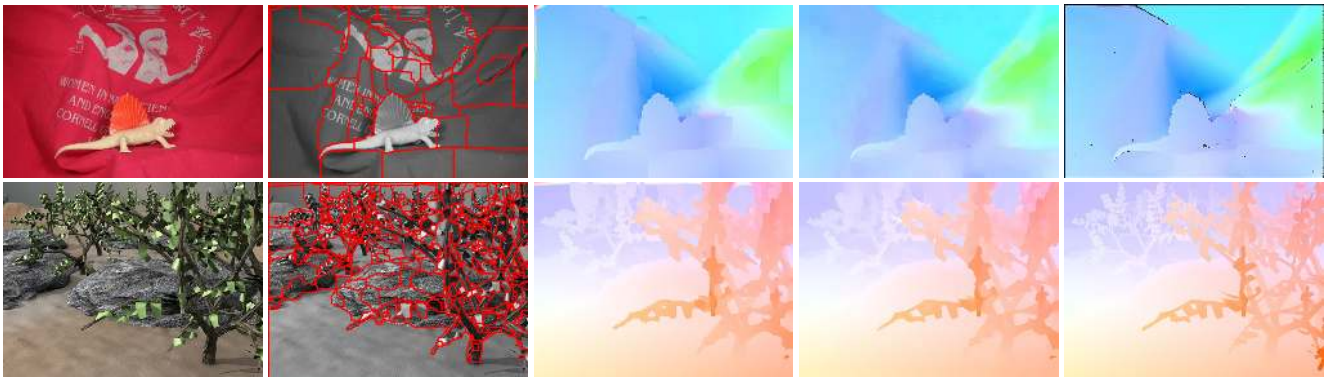
Figure 3: Effects of different MDL weights. Larger MDL weight leads to less homography models and larger pieces. We found that usually around $40 \sim 80$ homography models are adequate for flow estimation on KITTI benchmark.

Overlay of two input frames Estimated pieces Estimated flow (overlayed onto the image)

Figure 4: Example results of our method on KITTI benchmark. Note that in the example of the first row, motions of small objects such the road lamp and sign are successfully estimated (flow color scheme of the benchmark is used).

Table 2: Comparison with state-of-the-art two-frame optical flow methods on the *test* set of KITTI benchmark.

	>2 pixels		>3 pixels		>4 pixels		>5 pixels		End-Point	
	Out-Noc	Out-All	Out-Noc	Out-All	Out-Noc	Out-All	Out-Noc	Out-All	Avg-Noc	Avg-All
Our method	8.04 %	**13.76** %	**5.76** %	**10.57** %	**4.64** %	**8.84** %	**3.93** %	**7.72** %	**1.3 px**	2.9 px
NLTGV-SC [33]	**7.64** %	14.55 %	5.93 %	11.96 %	5.08 %	10.48 %	4.50 %	9.42 %	1.6 px	3.8 px
TGV2ADCSIFT [10]	8.04 %	17.87 %	6.20 %	15.15 %	5.24 %	13.43 %	4.60 %	12.17 %	1.5 px	4.5 px
BTF-ILLUM [18]	8.84 %	14.14 %	6.52 %	11.03 %	5.38 %	9.29 %	4.64 %	8.11 %	1.5 px	**2.8 px**
DeepFlow [43]	9.31 %	20.44 %	7.22 %	17.79 %	6.08 %	16.02 %	5.31 %	14.69 %	1.5 px	5.8 px
Classic+NL [37]	12.94 %	23.50 %	10.49 %	20.64 %	9.21 %	18.80 %	8.36 %	17.42 %	2.8 px	7.2 px
EPPM [2]	17.49 %	28.07 %	12.75 %	23.55 %	10.22 %	20.85 %	8.58 %	18.87 %	2.5 px	9.2 px
LDOF [12]	24.43 %	33.89 %	21.93 %	31.39 %	20.22 %	29.58 %	18.83 %	28.07 %	5.6 px	12.4 px



Reference frame Estimated pieces Flow w/o refinement Flow with refinement Ground-truth flow

Figure 5: Qualitative results on "Dimetrodon" and "Grove3" sequences of the Middlebury benchmark.

Table 3: Comparison of endpoint error on the *training* set of Middlebury benchmark.

	Dimetrodon	Grove2	Grove3	Hydrangea	RubberWhale	Urban2	Urban3	Venus
Ours	0.118	0.095	0.445	**0.146**	**0.072**	**0.196**	0.671	0.224
Ours w/o refinement	0.125	0.148	0.537	0.150	0.089	0.275	0.940	**0.190**
Classic+NL [37]	**0.115**	**0.091**	**0.438**	0.154	0.077	0.207	**0.377**	0.229
Hornáček *et al*. [21]	0.169	0.184	0.517	0.222	0.114	0.300	0.905	0.342

Table 4: Comparison of endpoint error with state-of-the-art methods on the *test* set of Middlebury benchmark. The numbers in brackets show the rank of each method on each sequence. Results of Unger *et al*. [40] are reproduced from their paper.

	Army	Mequon	Schefflera	Wooden	Grove	Urban	Yosemite	Teddy
Ours	0.08 (7)	0.21 (27)	0.23(9)	0.16 (30)	0.56 (7)	0.30 (5)	0.15 (54)	0.43 (8)
Classic+NL [37]	0.08 (7)	0.22 (33)	0.29 (25)	0.15 (19)	0.64 (18)	0.52 (48)	0.16 (65)	0.49 (19)
MDP-Flow2 [46]	0.08 (7)	0.15 (1)	0.20 (4)	0.15 (18)	0.63 (16)	0.26 (3)	0.11 (11)	0.38 (3)
NN-field [14]	0.08 (7)	0.17 (7)	0.19 (2)	0.09 (1)	0.41 (1)	0.52 (48)	0.13 (32)	0.35 (2)
Layer++ [38]	0.08 (7)	0.19 (15)	0.20 (4)	0.13 (6)	0.48 (3)	0.47 (36)	0.15 (54)	0.46 (13)
Unger *et al*. [40]	0.09	0.27	0.28	0.18	0.88	1.79	0.11	0.74

achieved. However the motion is complex, *e.g*. there are many non-rigid deformations, making it difficult for parametric model based methods.

We train the parameters on the *training* set, and set $\lambda_C = 0.5$, $\lambda_P = 2$, $\lambda_M = 100$, $\sigma_C = 100$ and $\theta = 1$. The MDL weight is tuned to be much smaller than that on the KITTI benchmark, as there are many complex motions and small objects, necessitating more homography models. Table 3 shows our results with and without refinement, compared to the Classic+NL method [37], and per-pixel homography estimation method [21]. In general, our final results are comparable to [37] on the *training* set. Compared to [21], without refinement our method outperforms [21] in 6 out of the 8 sequences, and outperforms it on all the sequences after refinement. Figure 5 shows some qualitative results of our method. Visually inspected, it gives smooth and accurate flow fields. It is able to group large regions with homogeneous (homography) motion (*e.g*. the ground and rocks in "Grove3"), meanwhile segment out the small regions with complex motions (*e.g*. the leaves). Figure 6 shows a challenging case (the "DogDance" sequence) with complex nonrigid motion. Our flow and segmentation are significantly better than [40], further demonstrating the ability of the proposed method in complex motion handling.

Table 4 compares the performance of our method versus others on the *test* set. As can be seen, our results are comparable to state-of-the-art methods. Note that our results are superior to or on par with results of [37] on all these sequences except for "Wooden", and outperform [40] on all the sequences except for "Yosemite".

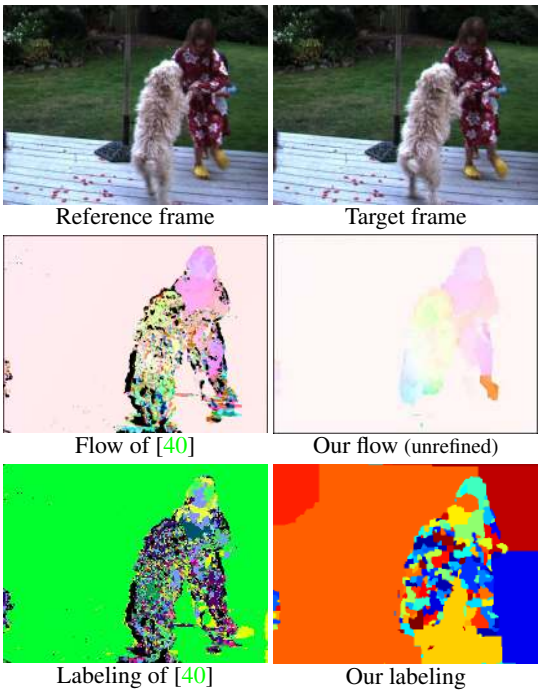


Figure 6: Comparison with [40] on the Middlebury "DogDance" sequence. Our method is less suffered from the complex non-rigid motion; the flow and labelling results are clear better than [40] (images reproduced from their paper).

5.3. Results on MPI Sintel

The Sintel benchmark contains long image sequences with large motions, severe illumination changes and specular reflections. Moreover, it contains large numbers of non-planar surfaces and complex non-rigid deformations, making it more challenging for the proposed piecewise parametric method.

We selected 23 image pairs (1 pair per sequence) from the clean pass of *training* set to train the parameter. After training we set $\lambda_C = 1$, $\lambda_P = 1$, $\lambda_M = 50$, $\sigma_C = 100$

Table 5: Comparison of end-point error with state-of-the-art methods on the *test* set of Sintel [13]. "all" / "noc" / "occ" indicate all / non-occluded / occluded regions respectively.

	Clean pass			Final pass		
	all	noc	occ	all	noc	occ
Our method	**4.388**	**1.714**	**26.202**	7.423	3.795	36.960
TF+OFM [24]	4.917	1.874	29.735	**6.727**	3.388	**33.929**
DeepFlow [43]	5.377	1.771	34.751	7.212	**3.336**	38.781
MDP-Flow2 [46]	5.837	1.869	38.158	8.445	4.150	43.430
EPPM [2]	6.494	2.675	37.632	8.377	4.286	41.695
Classic+NL [37]	7.961	3.770	42.079	9.153	4.814	44.509

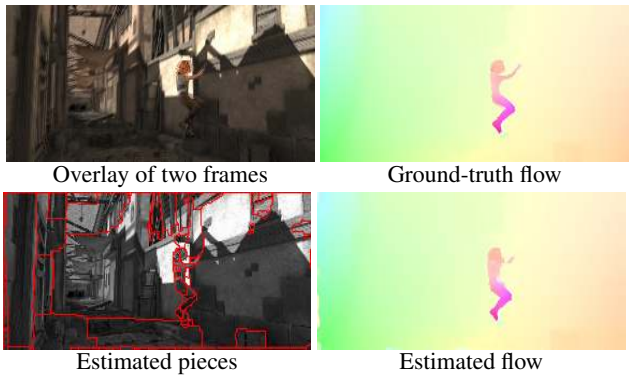


Figure 7: Sample results on the Sintel clean sequences.

and $\theta = 1.5$. The MDL is tuned to be very small due to the presence complex motions, *e.g.* the non-rigid motion of human and animal bodies. Figure 1 has shown a typical example and the result of our method, and an another example is presented in Fig. 7.

We then ran the method on the *test* set, and Tab. 5 presents the results of our method, compared with a few state-of-the-art methods. At the time of writing, our method ranks 2nd, and outperforms all published methods on the clean pass. Note that it performs especially well on the occluded regions, thanks to the use of parametric models in the post-processing stage. The proposed method performs inferiorly on the final pass, ranking 7th among all evaluated methods. We find that the synthetic atmospheric effects on the final sequences cause difficulties for both the PathMatch-based initialization, and our main algorithm. However, on the clean sequences for which the brightness constancy constraints satisfy, our method consistently produces accurate estimates.

5.4. Running time

The proposed method takes a few hundreds of seconds to estimate a forward flow field of size 640×480 in our experiment settings. The optimization time spent in Algorithm 1 is about $200 \sim 500$ seconds depending on the number of models. The initialization stage takes about 5 seconds and the refinement stage takes about 60 seconds.

6. Closing Remarks

We have presented a simple method for optical flow estimation using piecewise parametric model. Thanks to the new energy design and the joint discrete-continuous optimization, our method produces high-quality results that are superior to or comparable with state-of-the-art methods. We believe that piecewise parametric flow estimation deserves a position in highly accurate optical flow estimation.

Acknowledgement. This work was supported in part by the Natural Science Foundation of China (NSFC) under grant No. 61375044, the Specialized Fund for Joint Building Program of Beijing Municipal Education Commission, ARC Discovery grants DP120103896, DP130104567, and the ARC Centre of Excellence. HL thanks the support of NICTA Canberra Lab. JY is a CSC-co-funded BIT-ANU joint PhD student. The authors would like to thank AC and anonymous reviewers for invaluable comments.

References

[1] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision (IJCV)*, 92(1):1–31, 2011.

[2] L. Bao, Q. Yang, and H. Jin. Fast edge-preserving patchmatch for large displacement optical flow. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3534–3541, 2014.

[3] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (TOG)*, 28(3):24, 2009.

[4] P. Bhat, K. C. Zheng, N. Snavely, A. Agarwala, M. Agrawala, M. F. Cohen, and B. Curless. Piecewise image registration in the presence of multiple large motions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 2491–2497, 2006.

[5] S. Birchfield and C. Tomasi. Multiway cut for stereo and motion with slanted surfaces. In *International Conference on Computer Vision (ICCV)*, volume 1, pages 489–495, 1999.

[6] M. J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding (CVIU)*, 63(1):75–104, 1996.

[7] M. J. Black and A. D. Jepson. Estimating optical flow in segmented images using variable-order parametric models with local deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 18(10):972–986, 1996.

[8] M. Bleyer, C. Rhemann, and C. Rother. PatchMatch Stereo - Stereo matching with slanted support windows. In *British Machine Vision Conference (BMVC)*, pages 14.1–14.11, 2011.

[9] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 23(11):1222–1239, 2001.

[10] J. Braux-Zin, R. Dupont, and A. Bartoli. A general dense image matching framework combining direct and feature-based costs. In *International Conference on Computer Vision (ICCV)*, pages 185–192, 2013.

[11] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *European Conference on Computer Vision (ECCV)*, pages 25–36. 2004.

[12] T. Brox and J. Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(3):500–513, 2011.

[13] 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)*, pages 611–625, 2012.

[14] Z. Chen, H. Jin, Z. Lin, S. Cohen, and Y. Wu. Large displacement optical flow from nearest neighbor fields. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2443–2450, 2013.

[15] D. Cremers and S. Soatto. Motion competition: A variational approach to piecewise parametric motion segmentation. *International Journal of Computer Vision (IJCV)*, 62(3):249–265, 2005.

[16] A. Delong, A. Osokin, H. Isack, and Y. Boykov. Fast approximate energy minimization with label costs. In *CVPR 2010*, 2010.

[17] A. Delong, A. Osokin, H. N. Isack, and Y. Boykov. Fast approximate energy minimization with label costs. *International Journal on Computer Vision (IJCV)*, 96(1):1–27, 2012.

[18] O. Demetz, M. Stoll, S. Volz, J. Weickert, and A. Bruhn. Learning brightness transfer functions for the joint recovery of illumination changes and optical flow. In *European Conference on Computer Vision (ECCV)*, pages 455–471. 2014.

[19] A. A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. In *ACM SIGGRAPH*, 2001.

[20] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3354–3361, 2012.

[21] M. Hornáček, F. Besse, J. Kautz, A. Fitzgibbon, and C. Rother. Highly overparameterized optical flow using patchmatch belief propagation. In *European Conference on Computer Vision (ECCV)*, pages 220–234. 2014.

[22] H. Isack and Y. Boykov. Energy-based geometric multi-model fitting. *International Journal of Computer Vision (IJCV)*, 97(2):123–147, 2012.

[23] S. X. Ju, M. J. Black, and A. D. Jepson. Skin and bones: Multilayer, locally affine, optical flow and regularization with transparency. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 307–314, 1996.

[24] R. Kennedy and C. J. Taylor. Optical flow with geometric occlusion estimation and fusion of multiple frames. In *International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, pages 364–377, 2015.

[25] T. H. Kim, H. S. Lee, and K. M. Lee. Optical flow via locally adaptive fusion of complementary data costs. In *International Conference on Computer Vision (ICCV)*, pages 3344–3351, 2013.

[26] V. Kolmogorov and R. Zabin. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 26(2):147–159, 2004.

[27] C. Lei and Y.-H. Yang. Optical flow estimation on coarse-to-fine region-trees using discrete optimization. In *International Conference on Computer Vision (ICCV)*, pages 1562–1569, 2009.

[28] H. Li. Two-view motion segmentation from linear programming relaxation. In *CVPR 2007*, 2007.

[29] E. Mémin and P. Pérez. Hierarchical estimation and segmentation of dense motion fields. *International Journal of Computer Vision (IJCV)*, 46(2):129–155, 2002.

[30] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

[31] T. Nir, A. M. Bruckstein, and R. Kimmel. Over-parameterized variational optical flow. *International Journal of Computer Vision (IJCV)*, 76(2):205–216, 2008.

[32] C. Olsson and Y. Boykov. Curvature-based regularization for surface approximation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1576–1583, 2012.

[33] R. Ranftl, K. Bredies, and T. Pock. Non-local total generalized variation for optical flow estimation. In *European Conference on Computer Vision (ECCV)*, pages 439–454. 2014.

[34] C. Rother, L. Bordeaux, Y. Hamadi, and A. Blake. Autocollage. In *ACM Transactions on Graphics (TOG)*, volume 25, pages 847–852, 2006.

[35] A. Roussos, C. Russell, R. Garg, and L. Agapito. Dense multibody motion estimation and reconstruction from a handheld camera. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 31–40, 2012.

[36] C. Russell, J. Fayad, and L. Agapito. Energy based multiple model fitting for non-rigid structure from motion. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3009–3016, 2011.

[37] D. Sun, S. Roth, and M. J. Black. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision (IJCV)*, 106(2):115–137, 2014.

[38] D. Sun, E. B. Sudderth, and M. J. Black. Layered image motion with explicit occlusions, temporal consistency, and depth ordering. In *Annual Conference on Neural Information Processing Systems (NIPS)*, pages 2226–2234, 2010.

[39] D. Sun, E. B. Sudderth, and M. J. Black. Layered segmentation and optical flow estimation over time. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1768–1775, 2012.

[40] M. Unger, M. Werlberger, T. Pock, and H. Bischof. Joint motion estimation and segmentation of complex scenes with label costs and occlusion modeling. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1878–1885, 2012.

[41] C. Vogel, K. Schindler, and S. Roth. Piecewise rigid scene flow. In *International Conference on Computer Vision (ICCV)*, pages 1377–1384, 2013.

[42] J. Y. Wang and E. H. Adelson. Representing moving images with layers. *IEEE Transactions on Image Processing (TIP)*, 3(5):625–638, 1994.

[43] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid. DeepFlow: Large displacement optical flow with deep matching. In *International Conference on Computer Vision (ICCV)*, pages 1385–1392, 2013.

[44] J. Wills, S. Agarwal, and S. Belongie. What went where. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 37–44, 2003.

[45] L. Xu, J. Chen, and J. Jia. A segmentation based variational model for accurate optical flow estimation. In *European Conference on Computer Vision (ECCV)*, pages 671–684. 2008.

[46] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 34(9):1744–1757, 2012.

[47] K. Yamaguchi, D. McAllester, and R. Urtasun. Robust monocular epipolar flow estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1862–1869, 2013.

[48] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime TV-L^1 optical flow. In *Annual Pattern Recognition Symposium of the German Association for Pattern Recognition (DAGM)*, pages 214–223, 2007.

[49] C. Zhang, Z. Li, R. Cai, H. Chao, and Y. Rui. As-rigid-as-possible stereo under second order smoothness priors. In *European Conference on Computer Vision (ECCV)*, pages 112–126, 2014.

[50] C. L. Zitnick, N. Jojic, and S. B. Kang. Consistent segmentation for optical flow estimation. In *International Conference on Computer Vision (ICCV)*, volume 2, pages 1308–1315, 2005.