Dense, Accurate Optical Flow Estimation with Piecewise Parametric Model

Jiaolong Yang[1,2], Hongdong Li[2,3]
[1]Beijing Institute of Technology (BIT). [2]Australian National University (ANU). [3]ARC Centre of Excellence for Robotic Vision (ACRV)

This paper revisits the *piecewise parametric optical flow estimation* investigated in the 1990s [2]. We assume that the flow field can be piecewise represented by (possibly hundreds of) parametric motion models. To ease description, the 8-dof homography model is used. Our goal is to find suitable pieces (*e.g.* large ones for homogeneous motion and small ones for complex motion), and fit parametric models onto them for optical flow estimation. Compared to joint motion estimation and segmentation [4, 7], we regard the whole image as a single structure and piecewise approximate motion, rather than dividing it into multiple regions with independent motions.

Contributions. ***i)*** We represent and estimate the flow field with piecewise homography models, and solve the problem via joint discrete-continuous optimization. ***ii)*** We propose a novel energy formulation which takes into account both a piecewise constant model constraint, and a flow field continuity constraint. ***iii)*** We show that the piecewise parametric flow estimation used in the early days can be adapted to produce accurate results outperforming or on par with state-of-the-arts.

Formulation. Denote by $\mathbf{I}_1, \mathbf{I}_2$ the two images, and Ω the domain of $\mathbf{I}_1$. Let $\mathcal{L} = \{1,..,K\}$ be the labels of a homography set $\mathcal{H} = \{\mathrm{H}_k\}$, and $L : \Omega \to \mathcal{L}$ a labelling to generate pieces. Assigning label $k = L(\mathbf{x})$ to pixel $\mathbf{x}$ means that motion of $\mathbf{x}$ is induced by H_k. We define the energy function as

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

E_D* – *Data term. The data term E_D enforces the brightness/color constancy constraint using piecewise homography models as

$$E_D(\mathcal{H},L) = \sum_{\mathbf{x}\in\Omega} \rho\left((1-\alpha)|\mathbf{I}_1(\mathbf{x}) - \mathbf{I}_2(\mathrm{H}_{L(\mathbf{x})}\mathbf{x})| + \alpha|\nabla\mathbf{I}_1(\mathbf{x}) - \nabla\mathbf{I}_2(\mathrm{H}_{L(\mathbf{x})}\mathbf{x})|\right).$$

Note the abuse-of-notation: $\mathbf{x}$ and $\mathrm{H}\mathbf{x}$ represent inhomogeneous coordinates.

E_C* – *Flow continuity (inter-piece compatibility) term. This term enforces the continuity constraint of the flow field, rather than the widely used 1st-order or 2nd-order smoothness constraint (*e.g.* TV & TGV). Let $\mathcal{E}$ be the 4-connected neighbour set, E_C is defined to be

$$E_C(\mathcal{H},L) = \sum_{(\mathbf{x},\mathbf{x}')\in\mathcal{E}} w(\mathbf{x},\mathbf{x}') \cdot \rho\left(|\mathrm{H}_{L(\mathbf{x})}\bar{\mathbf{x}} - \mathrm{H}_{L(\mathbf{x}')}\bar{\mathbf{x}}|\right),$$

where $\bar{\mathbf{x}} = (\mathbf{x}+\mathbf{x}')/2$ is the midpoint of $(\mathbf{x},\mathbf{x}')$, and $w(\mathbf{x},\mathbf{x}')$ is color-based weight. The cost at $(\mathbf{x},\mathbf{x}')$ is nil if $L(\mathbf{x}) = L(\mathbf{x}')$, so E_C does not penalize the homography-induced motion variation within each piece even if it's large.

E_C enforces *inter-piece motion compatibility*; it allows compatible model switch, no matter how different the two models are. See top right images for an illustration of its effect.

E_P* – *Potts model term. We additionally use a pairwise Potts model term E_P to encourage spatially coherent labelling. This term is defined only on the discrete labelling variables as $E_P(L) = \sum_{(\mathbf{x},\mathbf{x}')\in\mathcal{E}} \delta\left(L(\mathbf{x}) \neq L(\mathbf{x}')\right)$, where $\delta(\cdot)$ is the 0-1 indicator function.

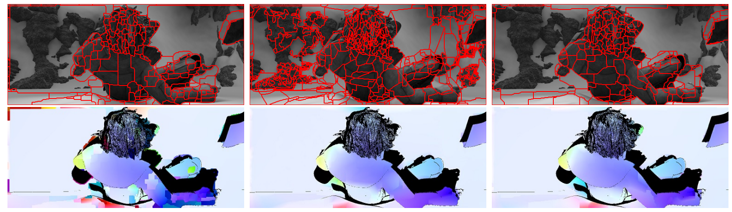
E_P enforces *intra-piece model constancy*; it penalizes any model switch, no matter how similar the two models are. See top right images for its effect.

E_M* – *MDL term. To reduce model redundancy, 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 $\tau(k) = \begin{cases} 1, & \text{if} \sum_{\mathbf{x}\in\Omega} \delta(L(\mathbf{x}) = k) > 0 \\ 0, & \text{othewise} \end{cases}$.

Optimization. We approach the joint discrete-continues problem via block coordinate descent that alternates between optimizing over L and $\mathcal{H}$.

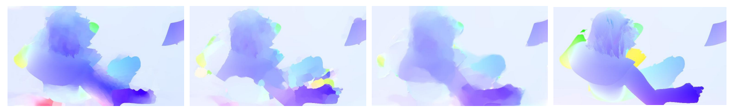
Solving for L with fixed $\mathcal{H}$ amounts to a labelling problem with multiple labels. The two piecewise terms E_C and E_P are sub-modular functions. We employ the graph-cut method [5] which can handel the label cost in E_M.

Solving for $\mathcal{H}$ with fixed L is an unconstrained continuous optimization problem. $\mathcal{H}$ appears only in E_D and E_C. As the models $\{\mathrm{H}_k\}$ interact with each other in E_C, we use an inner block coordinate decent to iteratively optimize $\{\mathrm{H}_k\}$ one by one, with the downhill simplex method.

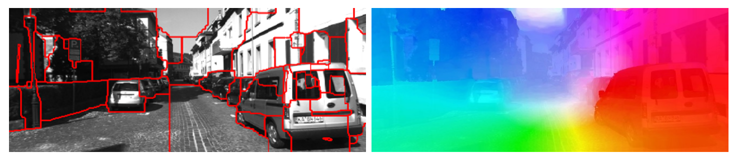


Without E_C Without E_P With E_C & E_P

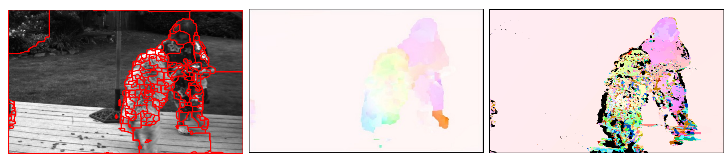
Left: noisy flow estimates for foreground complex motions.
Mid: noisy piece estimates for background homogeneous motions.
Right: good piece and flow estimates for all motions.



Estimated flow MDP-OF [8] LDOF [3] Ground truth



Estimated pieces Estimated flow



Estimated pieces Estimated flow (unrefined) Flow of [7]

We generate initial $\mathcal{H}$ and L from the nearest neighbour field [1]. Occlusion is detected via forward-backward consistency check. The Classic+NL-fast method [6] is used for flow refinement (on original image scale).

Performance. The method currently achieves leading performance on the KITTI benchmark (AEE 2.9), outperforms all published methods on the clean pass of the Sintel benchmark (AEE 4.4), and yields competitive results on the Middlebury benchmark (avg. rank 20.6). The alternation-based optimization takes about 200~500 seconds on a 640×480 image pair.

We believe that piecewise parametric flow estimation deserves a position in highly-accurate optical flow estimation, which is currently dominated by per-pixel translation flow estimation methods.

[1] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan Goldman. PatchMatch: A randomized correspondence algorithm for structural image editing. *SIGGRAPH*, 2009.

[2] Michael J Black and Allan D Jepson. Estimating optical flow in segmented images using variable-order parametric models with local deformations. *TPAMI*, 1996.

[3] Thomas Brox and Jitendra Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *TPAMI*, 2011.

[4] Daniel Cremers and Stefano Soatto. Motion competition: A variational approach to piecewise parametric motion segmentation. *IJCV*, 2005.

[5] Andrew Delong, Anton Osokin, Hossam N Isack, and Yuri Boykov. Fast approximate energy minimization with label costs. *IJCV*, 2012.

[6] Deqing Sun, Stefan Roth, and Michael J Black. Secrets of optical flow estimation and their principles. *CVPR*, 2010.

[7] Markus Unger, Manuel Werlberger, Thomas Pock, and Horst Bischof. Joint motion estimation and segmentation of complex scenes with label costs and occlusion modeling. *CVPR*, 2012.

[8] Li Xu, Jiaya Jia, and Yasuyuki Matsushita. Motion detail preserving optical flow estimation. *TPAMI*, 2012.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.