

Hierarchically-Constrained Optical Flow

Ryan Kennedy and Camillo J. Taylor
University of Pennsylvania

Optical flow algorithms attempt to estimate the perceived motion of each pixel between two frames of an image sequence [9]. Often, a grid-based graphical model that connects each pixel with its neighbors in the image and associates a motion vector with each image location. These motion vectors can either be treated as continuous or discrete. While optimization with continuous values allows for sub-pixel motion estimates [1], it can also result in convergence to local optima. This can be a particular problem for large motions, which are often missed in a coarse-to-fine local search.

In contrast, discrete graphical models can sometimes be solved optimally [8]. Even when the problems are NP-hard, approximate methods can often provide a bound on the globally-optimal energy [7]. These discrete optimization methods, however, tend to have computational complexities that are highly dependent on the size of the label space, and they cannot be readily applied to complex optical flow problems that may involve offsets of several hundred pixels between frames [2].

In this paper, we propose the use of a tree-based graphical model derived from a hierarchical image segmentation. In any choice of model, there is an inherent tradeoff between the model’s representational power and its optimization complexity. We argue that a hierarchical approach is capable of accurately modeling natural images while also being computationally tractable. Indeed, we show how the global optimum of this discrete optimization problem can be found by using efficient methods borrowed from the literature on deformable-parts models [5], even for problems involving hundreds of thousands of labels. This allows us to optimally solve large optical flow problems using a discrete, global model for the first time.

Our algorithm proceeds by constructing a tree-structured Markov random field (MRF) model based on a hierarchical segmentation of the first image (Figure 1). An example of a segmentation is shown in Figure 2. The goal of the motion estimation procedure is to assign an integral offset to each vertex that maps it onto its correspondent in the second image. This is done by defining a cost function that is small when pixels have a good correspondence and when each node has a similar offset to its parent.

We denote the image motions using a function $\mathbf{u} : \mathcal{V} \rightarrow \mathbb{Z}^2$. The cost function is defined over the graph structure for this displacement function \mathbf{u} :

$$\mathcal{C}(\mathbf{u}) = \lambda_0 \sum_{v \in \mathcal{V}} \mathcal{P}_v(\mathbf{u}(v)) + \lambda_1 \sum_{v \in \mathcal{V}} \mathcal{D}_v(\mathbf{u}(v)) + \sum_{(v_p, v_c) \in \mathcal{E}} \mathcal{S}_{v_p, v_c}(\mathbf{u}(v_p), \mathbf{u}(v_c)).$$

Here, $\mathcal{P}_v(\mathbf{u}(v))$ is a prior term that encourages each node to have a small offset. The term $\mathcal{D}_v(\mathbf{u}(v))$ is a unary matching term that measures how well vertex v is matched, and $\mathcal{S}_{v_p, v_c}(\mathbf{u}(v_p), \mathbf{u}(v_c))$ is a smoothness term defined over the set of edges in the hierarchical tree structure.

The goal of the optimization procedure is to find a solution which minimizes this cost function. Our model can be thought of as a large “deformable parts” model (DPM) [5], which is a widely-used framework in object recognition. The relationship between our model and a DPM allows us to leverage optimization techniques first developed in the DPM literature. First, because our graph is a tree, the minimum-energy solution can be found in polynomial time using a generalization of the Viterbi algorithm [4]. Also, because we use an $L1$ distance function, the cost matrices at each node can be computed very efficiently using a linear-time distance transform [3]. We also introduce several small approximations that dramatically speed up the algorithm while still yielding solutions that are very-nearly optimal.

Additionally, we describe a simple method of incorporating information from multiple frames in optical flow. We use the idea of *inertial estimates* from [6], where several estimates of the optical flow are computed using nearby frames and subsequently fused using a classifier. We show how the inertial estimates can instead be directly modeled in our cost function.

Experimentally, we show that state-of-the-art motion estimation schemes based on local optimization can have difficulties even on relatively simple

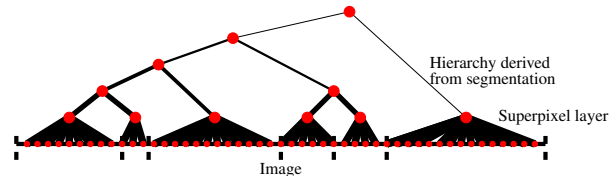


Figure 1: Depiction of our model, shown in 1D for simplicity. We use a hierarchical image segmentation, and each segment has an associated variable that is connected to its children. The root of the tree represents the entire image and leaves represent pixels. Edge weights are a function of the weights in the segmentation, denoted here by the thickness of black edges. Variables in the graphical model are denoted by red circles. After optimization, the final motion estimate is given by labels assigned to the pixel variables.

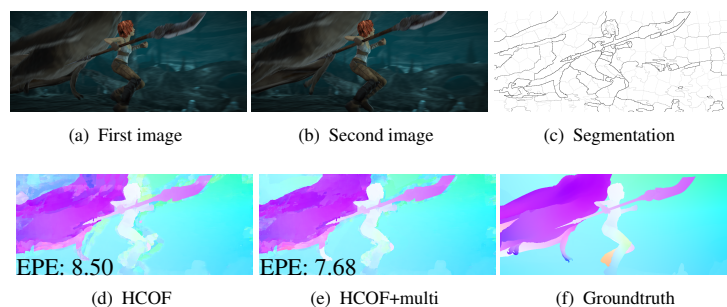


Figure 2: Result of our algorithms on an image from the Final dataset of MPI-Sintel. The flow images are colored with respect to the maximum groundtruth displacement.

motion analysis problems that contain large displacements. We illustrate this issue with both synthetic datasets and real images and show how our proposed global method can significantly improve performance in these situations. We also evaluate the proposed method on the challenging MPI-Sintel dataset [2] and compare its performance to other recent methods. In Figure 2, the results of our algorithm HCOF with and without the use of multiple inertial estimates are shown for an image from the MPI-Sintel dataset.

- [1] Thomas Brox, Andres Bruhn, Nils Papenberg, and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. *ECCV*, 2004.
- [2] Daniel J Butler, Jonas Wulff, Garrett B Stanley, and Michael J Black. A naturalistic open source movie for optical flow evaluation. In *ECCV*, pages 611–625. Springer, 2012.
- [3] Pedro Felzenszwalb and Daniel Huttenlocher. Distance transforms of sampled functions. Technical report, Cornell University, 2004.
- [4] Pedro F Felzenszwalb and Daniel P Huttenlocher. Pictorial structures for object recognition. *IJCV*, 61(1), 2005.
- [5] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *PAMI*, 32(9):1627–1645, 2010.
- [6] Ryan Kennedy and Camillo J. Taylor. Optical flow with geometric occlusion estimation and fusion of multiple frames. *EMMCVPR*, 2015.
- [7] Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(10):1568–1583, 2006.
- [8] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.
- [9] Deqing Sun, Stefan Roth, and Michael J Black. Secrets of optical flow estimation and their principles. *CVPR*, 2010.