This CVPR2015 extended abstract is the Open Access version, provided by the Computer Vision Foundation.

Learning to Detect Motion Boundaries

Philippe Weinzaepfel[a], Jerome Revaud[a], Zaid Harchaoui[a,b], Cordelia Schmid[a]
[a] LEAR team, Inria Grenoble Rhone-Alpes, Laboratoire Jean Kuntzmann, CNRS, Univ. Grenoble Alpes, France [b] NYU

Precise localization of motion boundaries is essential for the success of optical flow estimation, as motion boundaries correspond to discontinuities of the optical flow field. Furthermore, many computer vision tasks could benefit from the knowledge of accurate motion boundaries, *e.g.* action recognition, stereo depth computation, object segmentation in videos or object tracking. In this paper, we propose a learning-based approach for motion boundary detection, see Figure 1.

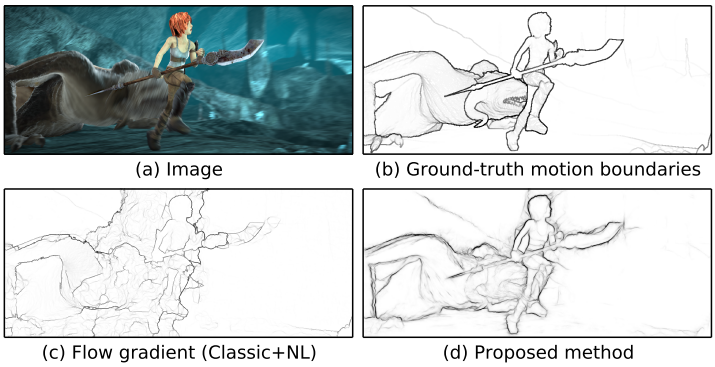

Figure 1: For the image in (a), we show in (b) its ground-truth motion boundaries, in (c) motion boundaries computed as the gradient of the Classic+NL flow [2], and in (d) our motion boundary detection. While also using the Classic+NL flow as a cue, our method is able to detect motion boundaries even in places where the flow estimation failed, like on the spear or the character's arm.

Method overview. We propose to extend the framework of structured random forests for edge detection in still images [1] to predict motion boundaries in videos. It relies on training several random trees that independently predict a binary motion boundary patch for each image patch, see Figure 2. The output of the different trees are then averaged to produce the final boundary detection. During training, we use images and ground-truth optical flows from the MPI-Sintel benchmark. Our approach leverages the following cues:

- appearance (RGB color),
- optical flow,
- optical flow error (*i.e.*, difference between the first image and the second image warped by the optical flow),
- backward optical flow and backward flow error.

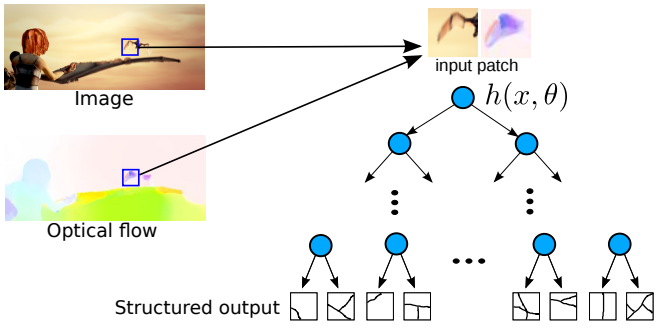

Figure 2: Illustration of the prediction process with our structured decision tree. Given an input patch from the left image (represented here by image and flow channels), we predict a binary boundary mask, *i.e.*, a leaf of the tree. Predicted masks are averaged across all trees and all overlapping patches to yield the final soft-response boundary map.

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

New dataset. We introduce the YouTube Motion Boundaries dataset (**YMB**), composed of 60 sequences taken from real-world videos with manually annotated motion boundaries. In contrast to MPI-Sintel and Middlebury, it comprises low-quality videos with important compression artifacts, and a larger diversity of scenes and characters.

Experimental results. We show that the proposed approach is both robust and computationally efficient. It significantly outperforms state-of-the-art motion-difference approaches on the MPI-Sintel, Middlebury and YMB datasets, see Figure 3. Moreover, we show that our approach is robust to failures in the optical flow estimation. We compare the results obtained with several state-of-the-art optical flow approaches and study the impact of the different cues used in the random forest.

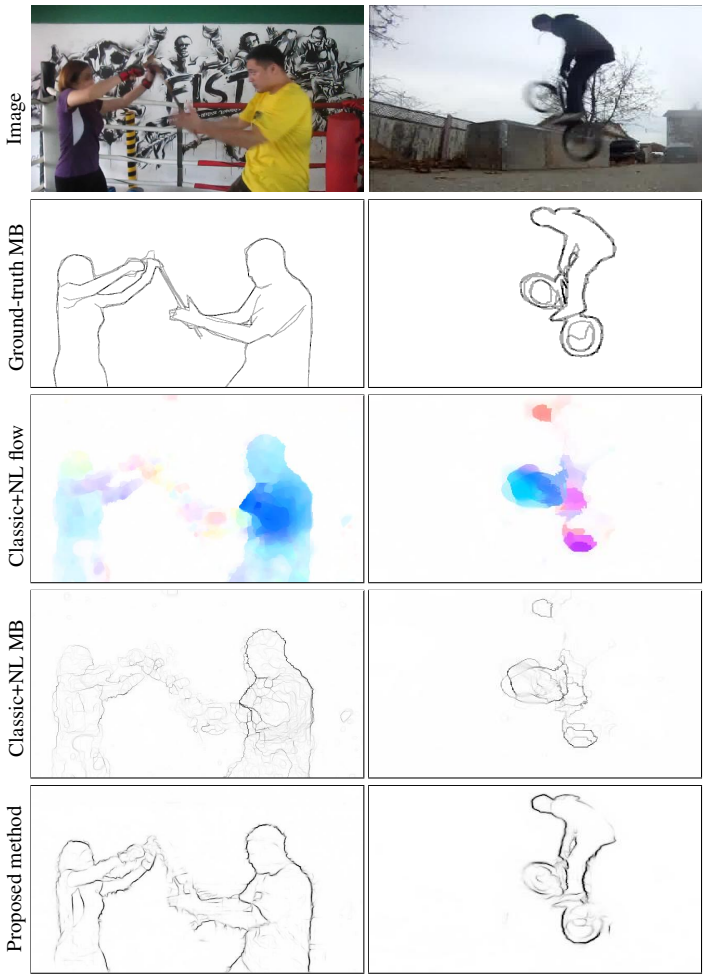

Figure 3: Example results for the YMB dataset with, from top to bottom: images, annotated motion boundaries, flow estimation using Classic+NL [2], norm of flow gradient, and the motion boundaries estimated by our method.

[1] P. Dollár and C. L. Zitnick. Structured forests for fast edge detection. In *ICCV*, 2013.
[2] D. Sun, S. Roth, and M. Black. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *IJCV*, 2014.