Towards Longer Long-Range Motion Trajectories

Michael Rubinstein[1]
mrub@mit.edu
Ce Liu[2]
celiu@microsoft.com
William T. Freeman[1]
billf@mit.edu

[1] MIT CSAIL
[2] Microsoft Research New England

Although dense, long-rage, motion trajectories are a prominent representation of motion in videos, there is still no good solution for constructing dense motion tracks in a truly long-rage fashion. Ideally, we would want every scene feature that appears in multiple, not necessarily contiguous, parts of the sequence to be associated with the same motion track. Despite this reasonable and clearly stated objective, there has been surprisingly little work on general-purpose algorithms that can accomplish that task. State-of-the-art dense motion trackers process the sequence incrementally in a frame-by-frame manner, and associate, by design, features that disappear and reappear in the video, with different tracks, thereby losing important information of the long-term motion signal.

In this paper, we propose a novel divide and conquer approach to long-range motion estimation. Given a long video or image sequence, we first produce high-accuracy *local* track estimates, or *tracklets*, and later propagate them into a *global* solution, while incorporating information from throughout the video. Tracklets are computed using state-of-the-art motion trackers [2, 3] that have become quite accurate for short sequences as demonstrated by standard evaluations. Our algorithm then constructs the long-range tracks by linking the short tracks in an optimal manner. This induces a combinatorial matching problem that we solve simultaneously for all tracklets in the sequence.

The main contributions of this paper are: (a) a novel divide-and-conquer style algorithm for constructing dense, long-rage motion tracks from a single monocular video, and (b) Novel criteria for evaluating long-range tracking results with and without ground-truth motion trajectory data. We evaluate our approach on a set of synthetic and natural videos, and explore the utilization of long-range tracks for action recognition.

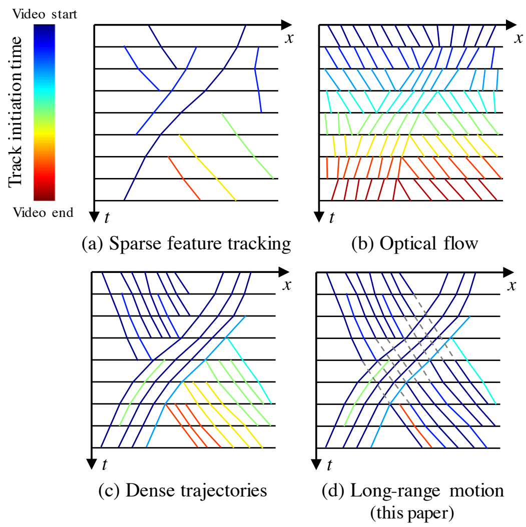

Figure 1: **Comparison of several motion representations.** Sparse feature point tracking (a) can establish long-range correspondences (*e.g.* between hundreds of frames), but only a few feature points are detected. While useful for some applications, it is a very incomplete representation of the motion in a scene. On the other hand, dense optical flow (b) reveals more about the moving objects, but the integer-grid-based flow fields cannot reliably propagate to faraway frames. A natural solution, therefore, is to combine feature point tracking and dense optical flow fields to a set of *spatially dense* and *temporally smooth* trajectories (or particles, tracks) [2], as show in (c). Despite recent advances in obtaining dense trajectories from a video sequence [1, 3], it is challenging to obtain *long-range* dense trajectories. In this paper, we propose a novel divide and conquer approach to long-range motion estimation (d), where point trajectories are assigned to the same tracks despite occlusion and deformation.

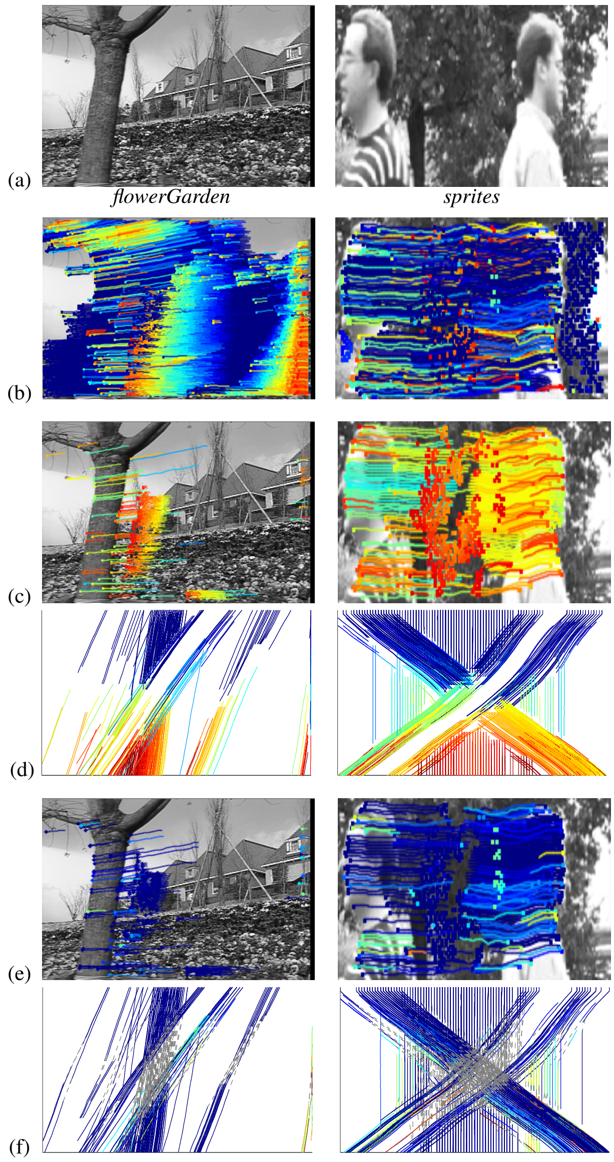

Figure 2: **Long-range motion trajectory results**. For each video (column), (a) is a representative frame from the sequence, (b) are the resulting long-range motion tracks, (c) and (e) focus on the tracks involved in the linkage (tracks which are left unchanged are not shown), before (c) and after (e) they are linked. (d) and (f) show XT views of the tracks in (c) and (e), respectively, when plotted within the 3D video volume (time advancing downwards). The tracks are colored according to their initiation time, from blue (earlier in the video), to red (later in the video). Track links are shown as dashed gray lines in the spatiotemporal plots (d) and (f). For clarity of the visualizations, random samples ($25 - 50\%$) of the tracks are shown.

[1] T. Brox and J. Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE transactions on pattern analysis and machine intelligence*, 33(3):500–513, 2011.

[2] P. Sand and S. Teller. Particle video: Long-range motion estimation using point trajectories. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2195–2202. IEEE, 2006.

[3] N. Sundaram, T. Brox, and K. Keutzer. Dense point trajectories by gpu-accelerated large displacement optical flow. *Computer Vision–ECCV 2010*, pages 438–451, 2010.