The Dense Long-Term Tracking Task

Applications like structure-from-motion or video editing benefit from long-term dense correspondences. Optical Flow gives dense correspondences, but only between pairs of frames.

**MFT: From two-frame optical flow to dense long-term trajectories**

Input: video. Output (for every frame): current position + visibility for each point from the first frame.

Baseline: Simple Chain of Flows

- Cannot recover from temporary occlusions
- Errors accumulate → drifting

x-axis: frame number
y-axis: point position (1D for clarity)

Baseline: Direct Flow From Template

- Does not accumulate errors, no drift
- Can recover after failure
- But harder task - change of viewpoint, illumination, large motion

MFT: Select the Most Reliable Chain

- Use the only option: Optical flow from template to frame #1
- Two correspondence candidates:
  - Direct match (top), Flow from frame #1 (bottom)
  - MFT selects one candidate (red), discards the rest
- Three correspondence candidates:
  - Direct match (bottom), from frame #1, from frame #2
- Select the most reliable chain at each frame, discard the other candidates.

Only consider a logarithmically spaced subset of candidates ($A=1, \Delta=2, A=4, \Delta=8, A=16, ...$) and the direct match ($A=\infty$)

Everything done independently for each reference frame pixel.

Uncertainty Estimation

- Select the candidate chain with the lowest uncertainty.
- Estimate flow vector uncertainty with a small CNN head, chain the uncertainties by summation.
- Trained on synthetic data using a standard uncertainty loss.

\[
L_u = \frac{1}{2} \sum \left( \frac{1}{2} \log \sigma^2 + \frac{1}{2} \left( \frac{\hat{u} - u}{\sigma} \right)^2 \right)
\]

Occlusion Estimation

- But do not select occluded chains.
- Optical flow methods trained to work in occluded regions (using context).
- Red arrow: flow correct, low uncertainty, but occlusion.
- Continued tracking would switch from the head to the bamboo.
- Trained on synthetic data using binary cross-entropy.

Optical Flow Adapted for MFT

- raw results, code, video
- optical flow CNN
- optical flow head
- uncertainty head
- occlusion head
- optical flow R x W
- uncertainty R x W
- occlusion R x W
- TRY MFT

Example MFT Application: Video Editing

A WOW! logo inserted in frame 0, color propagated by MFT to all frames.

Point-Tracking Benchmark Results

<table>
<thead>
<tr>
<th>Method</th>
<th>dense speed</th>
<th>sparse tracking quality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>acc / fps</td>
<td>true</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AP@50</td>
</tr>
<tr>
<td>TAPNet</td>
<td>5</td>
<td>35.0</td>
</tr>
<tr>
<td>PPS</td>
<td>500</td>
<td>0.002</td>
</tr>
<tr>
<td>OmniMotion</td>
<td>500</td>
<td>0.002</td>
</tr>
<tr>
<td>MFT (ours)</td>
<td>0.4</td>
<td>67.3</td>
</tr>
<tr>
<td>YAPSR</td>
<td>25</td>
<td>0.04</td>
</tr>
<tr>
<td>3CTracker</td>
<td>25</td>
<td>0.04</td>
</tr>
</tbody>
</table>

100+ FPS with pre-computed optical flows

MFT: good tracking quality, dense trajectories at high speed

Acknowledgments

This work was supported by Toyota Motor Europe, by the Grant Agency of the Czech Technical University in Prague, grant No.SGS22/173/0H3K3/37/13, and by the Research Center for Informatics project CZ.02.1.01/0.0/0.0/16_019/0000765 funded by OPVV.