1. Summary

We use path supervision to annotate a MOT dataset of unprecedented scale (720 sequences). The path trajectories were crowdsourced, since they are intuitive and natural to annotate. We hope it encourages and supports richer and fully data-driven MOT systems.

2. Annotation with Path Supervision

Most trajectory annotation frameworks interpolate through a set of manually annotated boxes, which are time-consuming and expensive to obtain. Instead, we annotate rough paths by loosely following the objects in the scene with the cursor in an intuitive manner.

Box-based annotation (LabelMe)

Path supervision

Paths represent a mode of weak-supervision on MOT trajectories. They provide no scale information and might be imprecise. So obtaining dense box-trajectories from them is not trivial.

Our energy minimization framework promotes the linkage of affine object detections while enforcing path constraints.

3. The PathTrack dataset

We use path supervision to annotate a MOT dataset of unprecedented scale (720 sequences). The path trajectories were crowdsourced, since they are intuitive and natural to annotate. We hope it encourages and supports richer and fully data-driven MOT systems.

Table 1: Comparison of PathTrack with other popular MOT datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th># seqs</th>
<th>Duration (mins)</th>
<th># tracks</th>
<th>Duration (mins)</th>
<th># tracks</th>
<th>Official MOT instances</th>
<th>Camera (Static/Moving)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KITTI15</td>
<td>21</td>
<td>13</td>
<td>50</td>
<td>30</td>
<td>-</td>
<td>-</td>
<td>car-mounted</td>
</tr>
<tr>
<td>MOT15</td>
<td>11</td>
<td>6</td>
<td>500</td>
<td>11</td>
<td>10</td>
<td>1221</td>
<td>P</td>
</tr>
<tr>
<td>MOT16</td>
<td>7</td>
<td>4</td>
<td>121</td>
<td>4</td>
<td>4</td>
<td>385</td>
<td>S+M</td>
</tr>
<tr>
<td>PathTrack (ours)</td>
<td>640</td>
<td>161</td>
<td>15,380</td>
<td>80</td>
<td>11</td>
<td>172</td>
<td>P</td>
</tr>
<tr>
<td>VirtualKITTI [1]</td>
<td>640</td>
<td>161</td>
<td>15,380</td>
<td>80</td>
<td>11</td>
<td>172</td>
<td>P</td>
</tr>
<tr>
<td>MOT15 [3]</td>
<td>7</td>
<td>4</td>
<td>121</td>
<td>4</td>
<td>4</td>
<td>385</td>
<td>S+M</td>
</tr>
<tr>
<td>MOT16 [4]</td>
<td>7</td>
<td>4</td>
<td>121</td>
<td>4</td>
<td>4</td>
<td>385</td>
<td>S+M</td>
</tr>
</tbody>
</table>

Apart from quantity, diversity is a central goal of the dataset. It includes 7 different kind of scenes taken from both stationary and moving cameras. Each sequence is labeled with its scene-type and camera-movement, allowing for fine-grained performance analysis.

4. Viewpoint

Path supervision (ours)

Box-based annotation (LabelMe)

Annotate while watching

Path supervision (ours)

Annotating trajectories is expensive!

Annotation efficiency with path supervision

Annotation time (hours)

The community can still benefit from even more training data for the matching problem. It can be efficiently collected with path supervision, e.g., the MOT15 data can be annotated 20 times faster while achieving the same matching accuracy.

5. Experiments

5.1 Tracking results

First tracking results on the dataset demonstrate the usefulness of our training data.

Matching accuracy (%)

Matching accuracy (%)

The full dataset can be downloaded from: http://people.ee.ethz.ch/~smanenfr/pathtrack/