IPT: A Dataset for Identity Preserved Tracking in Closed Domains

Thomas Heitzinger
theitzinger@cvl.tuwien.ac.at

Martin Kampel
martin.kampel@tuwien.ac.at
Motivation

• The IPT dataset aims to facilitate the application of state-of-the-art computer vision methods to privacy-critical applications such as:
  • Ambient Assisted Living (AAL)
  • Security

• Privacy is protected by using a depth sensor as an anonymizing imaging method. In many cases behavior analysis does not need to know who is acting but rather what they are doing.
Contribution

- A new public dataset for identity preserved human detection and tracking with flexible labelling for use in either 2d or 3d space.

- Baseline results for person detection and tracking.
Dataset Overview

- 10 sequences featuring:
  - Indoor environments
  - Both scripted and natural behavior
  - Frequent person-person and person-object occlusions

- 72k frames in total, static depth sensor at 640×480 resolution and 30 fps.

- Split by sequence into training (70%), validation (20%) and test sets (10%).
Dataset Labeling

• 3d location of actors
  • Consistent actor ID across all sequences

• 2d bounding boxes
  • Inferred from 3d locations

• Semantic room layout layers
  • Exterior space (walls, outside space)
  • Obstructed space (furniture, appliances, ...)
  • Semi-obstructed space (chairs, beds, ...)

• Sensor pose

• Tracking ground truth in the MOT Challenge format
Baseline Object Detection

• 2d detection baseline using a YOLOv3 [1] model modified for single person class output.

• Model performance is improved using an efficient background model [2].
  • Three channel input \((x_t, \mu_t, \sigma_t)\) with depth frames \(x_t\) and geometrically weighted mean \(\mu_t\) and standard deviation \(\sigma_t\).
  • Updated once per second.
  • During training an approximation of the background model is computed using a ten second window preceding each sample.

Object Detection Results

- Results show significant variation in model performance across sequences due to different room layouts, sizes and occlusion profiles.
- Comparison of validation set and test set performance suggests non-representative test set, addressed in upcoming work.
- YOLOv3 Tiny included for potential use in target applications with inference on embedded devices.

<table>
<thead>
<tr>
<th></th>
<th>Average Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Validation</td>
</tr>
<tr>
<td>YOLOv3</td>
<td>77.2%</td>
</tr>
<tr>
<td>YOLOv3 + BG Model</td>
<td>85.2%</td>
</tr>
<tr>
<td>YOLOv3 Tiny</td>
<td>71.9%</td>
</tr>
</tbody>
</table>
Baseline Tracking

- Evaluation using full MOT Challenge metrics (CVPR19 benchmark).

<table>
<thead>
<tr>
<th>Mode</th>
<th>MOT Accuracy</th>
<th>Identity F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SORT</td>
<td>76.6%</td>
<td>14.5%</td>
</tr>
<tr>
<td>DeepSORT</td>
<td>77.2%</td>
<td>17.5%</td>
</tr>
<tr>
<td>DeepSORT Short Tracks</td>
<td>77.3%</td>
<td>58.8%</td>
</tr>
</tbody>
</table>

- Results suggest utility of deep features for reidentification even for pure depth data.
- “Short Tracks” change track ID whenever subjects leave the sensors’ field of view.

Thank you