**Position**

- Human Pose Estimation in the Wild has great opportunities in animation, action recognition, intention recognition and prediction for autonomous driving
- Current State-of-the-art focuses on RGB lacking global precision
- LiDAR sensors provide centimeter precision
- Human Pose Estimation using both RGB and LiDAR provides high precision and enables new use cases

**3D Pose Estimation**

**Task:**
- Prediction of the joints of a human in 3D Space
- A point \(x = (x, y, z)^T\) describes the position of the joint
- Most approaches only predict relative to the position of the hip of the person. In contrast to them we predict the 3D positions relative to the observer.

**Methodologies:**
- **Bottom Up:** Predict the position of the joints in heatmaps and then construct skeletons for the persons from the joints.
- **Top Down:** First find the persons in the scene, then predict the joint positions per person.
  - LCR-Net[3,4] with its Faster-RCNN[1]-like architecture was most influential to our work
  - First an RPN predicts 2D Boxes, then a refinement predicts deltas to 3D pose anchors

**Limitations of RGB:**
- **Depth Ambiguity:** In an RGB image, a large person (blue) further away than a smaller person (pink) can occupy the same pixel space. Given only the image a correct depth estimation proves difficult.

**Sensor Fusion RGB and LiDAR in 3D Detection**

Importance of LiDAR
- RGB-only approaches lack behind LiDAR + RGB approaches
- Best Detectors on KITTI are mostly LiDAR + RGB Fusion approaches
- LRPD [5] shows characteristics of fusion schemas for pedestrians

Reference Fusion Schema
- AVOID [2] is a Faster-RCNN[1]-like two stage approach
- First ROI crop using the anchors, fuse and predict region proposals
- Secondly ROI crop using proposals, fuse and predict the final boxes
- This simple schema can be adapted to other tasks as well

**HPERL Architecture**

- Two Stage Approach (Faster-RCNN[1]-like)
  - **1st Stage:**
    - Fusion of LiDAR and RGB
    - ROI align operation, then averaged or concatenated
  - **2nd Stage:**
    - **Pose Refinement:**
      - Predict deltas to add onto the absolute anchor poses generated
    - **Anchor Scoring:**
      - Estimate scores for each absolute anchor pose

**Sensor Fusion in Human Pose Estimation**

- Improve pose estimation by using LiDAR and RGB

**Advantages:**
- **Localization:** Using LiDAR enables high localization precision
- **Occlusion:** Combination of LiDAR and RGB helps in occlusion cases, as the LiDAR with the elevated position suffers less from occlusion

**Precise Pose:**
- Using multiple sensor modalities allows for the precision in pose estimation of RGB and precise 3D understanding with LiDAR

**Evidence:**
- PedX Pose Estimation Dataset (largest best: PCKh, smallest: else)

<table>
<thead>
<tr>
<th>Method</th>
<th>Type</th>
<th>2D MPJPE</th>
<th>PCKh</th>
<th>Center Dist. Err.</th>
<th>XY Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGB-Baseline</td>
<td>2D</td>
<td>87.76px</td>
<td>65.02%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>3D</td>
<td>87.66px</td>
<td>65.92%</td>
<td>4.88m</td>
<td>1.44m</td>
</tr>
<tr>
<td>Fusion</td>
<td>2D</td>
<td>45.66px</td>
<td>70.08%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>3D</td>
<td>45.65px</td>
<td>70.22%</td>
<td>0.95m</td>
<td>0.39m</td>
</tr>
</tbody>
</table>

**References**


**Contact Information**

Michael Fürst
michael.fuerst@dfki.de
http://av.dfki.de/members/fuerst/

René Schuster
rene.schuster@dfki.de
http://av.dfki.de/members/schuster/