Temporal Segmentation of Egocentric Videos

VIVEK PRADHAN
Problem Statement

(a) Car  (b) Bus  (c) Walking  (d) Sitting

(e) Wheels  (f) Standing  (g) Static
DataSet
DataSet

- Show External Video.
Problem Statement

- Input Video
  - Stationary
    - Static
      - Sitting
    - Moving Head
      - Standing
  - Transit
    - Open View
      - Walking
    - Box
      - Car
      - Bus
Problems with Short Term motion detectors
Problems with Short Term motion detectors
Cumulative displacement curves

- Divide Image into 10x5 cells and compute displacement of cells.
- Make it cumulative over time to cancel out head motions.
Properties of CD curves
Properties of CD curves
Properties of CD curves

- Sitting
- Walking
- Riding Bus
SVM features from CD curves.

- Radial Projection Response
- Motion Clusters
- FOE estimation
Gaze Fixation through CD curves

- Causes head motion to stop.
- This can be exploited to find the gaze fixation.
Experiments and Results

<table>
<thead>
<tr>
<th></th>
<th>Walking</th>
<th>Car</th>
<th>Standing</th>
<th>Bus</th>
<th>Wheels</th>
<th>Sitting</th>
<th>Static</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking</td>
<td>83%</td>
<td>0%</td>
<td>6%</td>
<td>6%</td>
<td>4%</td>
<td>1%</td>
<td>0%</td>
</tr>
<tr>
<td>Car</td>
<td>1%</td>
<td>74%</td>
<td>3%</td>
<td>15%</td>
<td>0%</td>
<td>3%</td>
<td>4%</td>
</tr>
<tr>
<td>Standing</td>
<td>14%</td>
<td>2%</td>
<td>47%</td>
<td>4%</td>
<td>0%</td>
<td>31%</td>
<td>2%</td>
</tr>
<tr>
<td>Bus</td>
<td>3%</td>
<td>19%</td>
<td>27%</td>
<td>43%</td>
<td>0%</td>
<td>7%</td>
<td>1%</td>
</tr>
<tr>
<td>Wheels</td>
<td>9%</td>
<td>0%</td>
<td>0%</td>
<td>6%</td>
<td>86%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Sitting</td>
<td>3%</td>
<td>1%</td>
<td>22%</td>
<td>1%</td>
<td>0%</td>
<td>62%</td>
<td>10%</td>
</tr>
<tr>
<td>Static</td>
<td>0%</td>
<td>1%</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>97%</td>
</tr>
</tbody>
</table>

Table 2: Confusion matrix for the cascaded classifier tree. Rows are ground truth. Diagonal elements represent class accuracy, off diagonal elements give pairwise confusion.
# Experiments and Results

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy</th>
<th># Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg.</td>
<td>Class 1</td>
</tr>
<tr>
<td>Box vs. Open</td>
<td>93%</td>
<td>94%</td>
</tr>
<tr>
<td>Car vs. Bus</td>
<td>74%</td>
<td>73%</td>
</tr>
<tr>
<td>Sitting vs. Standing</td>
<td>70%</td>
<td>72%</td>
</tr>
<tr>
<td>Static vs. Moving</td>
<td>96%</td>
<td>98%</td>
</tr>
<tr>
<td>Stationary vs. Transit</td>
<td>90%</td>
<td>86%</td>
</tr>
<tr>
<td>Walking vs. Wheels</td>
<td>93%</td>
<td>96%</td>
</tr>
</tbody>
</table>
CD curves for other classes and tasks

- Might have very different results for chest mounted cameras.
- Authors have performed the same experiment with 3D CNNs.
- Activity recognition.