Alejandro Cartas, Petia Radeva, and Mariella Dimiccoli

Modeling Long-Term Interactions to Enhance Action Recognition

Approach

1. Primary action region: Hands
2. Secondary action regions

Motivation and goal

Comparative Results

Key contribution

Datasets

No motion features involved

Egocentric contextual cues (hands+objects)

Gaze evolution

Egocentric contextual cues (hands+objects)

Action & activities have spatio-temporal context:

1. Spread peanut on bread
2. Take knife
3. Pour milk on cup
4. Scoop peanut with knife
5. Put plastic spatula
6. Open bread container
7. Open Fridge
8. Cut tomato with knife
9. Turn on burner
10. Take knife to plate
11. Put tomato on plate
12. Take plastic spatula
13. Make Cheese Sandwich
14. Pour Mayonnaise
15. Close Mayonnaise

Primary Regions

Secondary Regions

Secondary action regions

Up to 2K regions are proposed as candidates. A region is chosen after classification. If 10 to 2K regions are available, the temporal evolution of the region is classified as an action or activity. The classification is made according to the motion features detected in the region. If the motion features detected are consistent with the action or activity, the region is classified as an action or activity. Otherwise, the region is classified as noise.

True Prediction Examples

False Prediction Examples

<table>
<thead>
<tr>
<th>Datasets</th>
<th>#Actions</th>
<th>#People</th>
<th>#Activities</th>
<th>#Videos</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTEA+</td>
<td>61 &amp; 71</td>
<td>40</td>
<td>37</td>
<td>4</td>
</tr>
<tr>
<td>(Li et al, 2011)</td>
<td>25 &amp; 40</td>
<td>14</td>
<td>25 &amp; 40</td>
<td>2</td>
</tr>
<tr>
<td>(Fathi et al, 2012)</td>
<td>7</td>
<td>7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(Comaniciu and Lang, 2012)</td>
<td>4</td>
<td>4</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Datasets</th>
<th>#Activities</th>
<th>#People</th>
<th>#Actions</th>
<th>#Videos</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTEA</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>(Li et al, 2011)</td>
<td>7</td>
<td>7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(Fathi et al, 2012)</td>
<td>4</td>
<td>4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(Comaniciu and Lang, 2012)</td>
<td>25 &amp; 40</td>
<td>14</td>
<td>25 &amp; 40</td>
<td>2</td>
</tr>
</tbody>
</table>