UPC at MediaEval 2013: Hyperlinking Task

Carles Ventura
Marcel Tella-Amo
Xavier Giró-i-Nieto

Challenge: Visual Features only

SURF descriptors
[Bay et al, CVIU 2008]

Bag of Features
[Sivic & Zisserman, 2003]

Histogram intersection

Figure: Fergus, ICCV 2009

Approach

Query video

Shot boundaries

Context

Anchor

Context

Keyframe ranked lists

Fused ranked list

Linked segments

Results

<table>
<thead>
<tr>
<th></th>
<th>MAP</th>
<th>P@5</th>
<th>P@10</th>
<th>P@20</th>
</tr>
</thead>
<tbody>
<tr>
<td>No context</td>
<td>0.0282</td>
<td>0.2600</td>
<td>0.2000</td>
<td>0.1233</td>
</tr>
<tr>
<td>Context</td>
<td>0.0260</td>
<td>0.2400</td>
<td>0.1967</td>
<td>0.1217</td>
</tr>
</tbody>
</table>

Conclusions

- Visual alone is useful, but not enough.
- 1.2M keyframes is challenging, video is more.
- Performance can be improved with spatial coding, larger vocabulary and concept detectors.