DRESS LIKE A STAR: RETRIEVING FASHION PRODUCTS FROM VIDEOS

NOA GARCIA & GEORGE VOGIATZIS

(Context)

To find fashion items in videos using captures from scenes.

Challenges

- Average movie duration: 120 minutes
- Standard FPS rate: 24 fps
- Average frames per movie: 172,800 frames

Our System

1. Fashion items and frames are related in a database.
2. BRIEF features are extracted, tracked and aggregated into key features, which are indexed in a kd-tree.
3. Queries are used to find the frame and retrieve fashion products associated with it.

Constraints

- Instead of object recognition, system based on frame retrieval.

Training Phase

Why binary?

Kd-Tree Indexing

Key Features using majorities

Query Phase

Frames within the most voted shot

Experiments

Query images are captured during video playback. The frame number of the acquired image is annotated as its ground truth.

Retrieval Performance

Similar accuracy, huge gain in memory requirements.

<table>
<thead>
<tr>
<th>Indexed Features</th>
<th>BF</th>
<th>KT</th>
<th>KF</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory</td>
<td>2.53GB</td>
<td>2.53GB</td>
<td>762MB</td>
<td>61MB</td>
</tr>
<tr>
<td>B = 10</td>
<td>0.90</td>
<td>0.94</td>
<td>0.92</td>
<td>0.93</td>
</tr>
<tr>
<td>B = 100</td>
<td>0.93</td>
<td>0.94</td>
<td>0.93</td>
<td>0.94</td>
</tr>
<tr>
<td>B = 250</td>
<td>0.97</td>
<td>0.93</td>
<td>0.94</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Scaleability

- 40 movies
- 80 hours
- 7 million frames