Video retrieval by mimicking human poses

J. Nataraj¹, A. Zisserman², M. Eichner¹, V. Ferrari³ and C. V. Jawahar¹
¹CVIT, IIIT Hyderabad
²Visual Geometry Group, University of Oxford
³Calvin Group, ETH Zurich

1. Objective

- New access mechanism for retrieving videos
- Fast, accurate and large scale retrieval
- Interactive and intuitive interface for the user

2. Pose search Overview

Process videos (Offline)

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eichner et al. [1]</td>
<td>9.1</td>
<td>12.1</td>
</tr>
<tr>
<td>Intersection, PCP = 1</td>
<td>41.1</td>
<td>3.5</td>
</tr>
<tr>
<td>Intersection, PCP &gt; 0.83</td>
<td>16.7</td>
<td>5.1</td>
</tr>
<tr>
<td>Intersection, PCP &gt; 0.87</td>
<td>7.8</td>
<td>5.5</td>
</tr>
</tbody>
</table>

3. Pose representation and matching

\[ \theta = \cos(\theta_A \cdot \sin(\theta_B)) \]

Observation

- Angular difference between parts distinguishes poses well.
- Negative cosine of angular difference monotonically increases.

Representation

- The proposed representation of pose A for each part i is
  \[ \theta_i = \{ \cos(\theta_i), \sin(\theta_i) \} \]
- For six upper body parts, the pose representation is 12-dimensions.

Matching

- The dissimilarity between pose representations of A and B are
  \[ (r_i - r_B)^2 = \sum_2 \{ 1 - \cos(\theta_i - \theta_B) \} \]

4. Fast, accurate and large scale retrieval

Accuracy

- Filter poses based on consensus of two estimation methods
  - Run [1,2] on a frame and find corresponding poses using upper body detector [1]
  - Measure agreement of poses using PCP. Accept if above threshold

Speed

- Randomized K-D retrieves K-NN approximately with high recall rate.
- Combined with our compact representation, it retrieves at high speed.

Large database

- 3.1 million frames from 22 movies.
- 1.8 million pose estimates detected by the algorithms.
- 50K pose estimates after combining the pose estimates from [1,2] (threshold: PCP=1).

5. Engaging interfaces

- Upload an image of a person.
- The system takes the best detected pose as query.
- Move the interactive stickman.
- Strike a pose before Kinect.
- The system responds to non-trivial movements of the user.

6. Video retrieval system

Performance

- *MAP

Video Retrievals

Contributions

- Pose combination methods for improved accuracy.
- Compact but descriptive pose representations.
- Fast and large scale retrieval.
- Intuitive and entertaining interface.