3D-HOG EMBEDDING FRAMEWORKS FOR SINGLE AND MULTI-VIEWPOINTS ACTION RECOGNITION BASED ON HUMAN SILHOUETTES

Federico Angelini*, Zeyu Fu*, Sergio A. Velastin†, Jonathon A. Chambers*, Syed Mohsen Naqvi*

*Intelligent Sensing and Communications Research Group, Newcastle University, UK
†Department of Computer Science and Engineering, University Carlos III de Madrid, Spain

1 Overview

Given the high demand for automated systems for human action recognition, great efforts have been undertaken in recent decades to progress the field [1]. In this paper, we present frameworks for single and multi-viewpoints action recognition based on:

- Space-Time Volume (STV) of human silhouettes
- 3D-Histogram of Oriented Gradient (3D-HOG) Embedding [2]

Our contributions

- 3D-HOG Embedding [2] based frameworks exploiting local gestures analysis
- Single and multi-viewpoints cases
- Accuracy and robustness to appearance changes
- Outperforming results on Weizmann and i3DPost datasets

2 Baseline Method

- Baseline method: 3D-HOG Embedding [2]
- It defines the basic data processing structure (Fig 2), also used in the Proposed Frameworks

Key drawbacks

- Attention problem (Fig 3): it has not been addressed;
- Performance stability: affected by randomly selected library in the Embedding phase;
- Action-labels-based local classifiers, without considering cross-location local gestures relationships;

3 Proposed Frameworks

Main ideas

- Locally, actions look like simpler gestures
- Globally, an action can be seen as a particular combination of local gestures (Fig 4)

Hierarchical Clustering for Gestures Library

\[ (b_{ij})^n_{m} = (b_{i,j})_{1}^1 \cup (b_{i,j})_{1}^2 \cup \ldots \cup (b_{i,j})_{1}^{nLW} \]

Multi-viewpoints framework

Fig 4. Example of local gestures analysis for peculiar gestures composition. Letters represent gesture labels.

4 Results

Comparison between baseline and proposed frameworks performance

Experimental Setting: Leave-one-actor-out (robustness to appearance changes)

<table>
<thead>
<tr>
<th>Method</th>
<th>Actions</th>
<th>Accuracy (%)</th>
<th>B</th>
<th>A1</th>
<th>A2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed Framework</td>
<td>10</td>
<td>100%</td>
<td>20</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Proposed Framework</td>
<td>8</td>
<td>99.60%</td>
<td>30</td>
<td>95%</td>
<td></td>
</tr>
</tbody>
</table>

Conclusions

- Outperforming results in all studied cases
- Stable performance over different trainings
- Higher accuracy for smaller α (best values)

Comparison between dataset state-of-arts and proposed frameworks performance

Experimental Setting: Leave-one-actor-out (robustness to appearance changes)

<table>
<thead>
<tr>
<th>Method</th>
<th>Actions</th>
<th>Accuracy (%)</th>
<th>B</th>
<th>A1</th>
<th>A2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed Framework</td>
<td>10</td>
<td>100%</td>
<td>20</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Proposed Framework</td>
<td>8</td>
<td>99.73%</td>
<td>30</td>
<td>99%</td>
<td></td>
</tr>
</tbody>
</table>

Conclusions

- State-of-art results (Weizmann) and outperforming results (i3DPost)

Examples

- True Label: WALK
- Estimated Label: HWAV
- Accuracy: 95%

5 References