Large-scale Continuous Gesture Recognition Using Convolutional Neutral Networks

August 13, 2016

1 Team details

- Team name: AMRL
- Team leader name: Pichao Wang
- Team leader address, phone number and email: 3/68, Robsons Road, Wollongong, Australia, (+61)405278871, pw212@uowmail.edu.au
- Rest of the team members: Wanqing Li, Song Liu, Yuyao Zhang, Zhimin Gao and Philip Ogunbona
- Affiliation: University of Wollongong

2 Contribution details

- Title of the contribution: Large-scale Continuous Gesture Recognition Using Convolutional Neutral Networks
- Final score: 0.2655
- General method description: This paper addresses the problem of continuous gesture recognition with convolutional neural networks (ConvNets) using depth maps sequences. Unlike the common isolated recognition scenario, the gesture boundaries are here unknown, and one has to solve two problems: segmentation and recognition. For segmentation, we first obtained the begin and end frames of each gesture based on quantity of movement (QOM) and then proposed one compact representations for depth sequences, called Improved Depth Motion Map (IDMM), which converts each depth sequence into one image, to recognize the gestures using ConvNets. This method enables the use of existing ConvNets models directly on video data with fine-tuning, without introducing much parameters to be learned.
- References:
References


- Representative image / diagram of the method:

![Figure 1: The framework for proposed method.](image)

- Describe data preprocessing techniques applied (if any): None
3 Visual Analysis

3.1 Gesture Recognition (or/and Spotting) Stage

3.1.1 Features / Data representation
Describe features used or data representation model FOR GESTURE RECOGNITION (OR/AND SPOTTING) STAGE (if any): Deep learned features using ConvNets.

3.1.2 Dimensionality reduction
Dimensionality reduction technique applied FOR GESTURE RECOGNITION (OR/AND SPOTTING) STAGE (if any): None

3.1.3 Compositional model
Compositional model used, i.e. pictorial structure FOR GESTURE RECOGNITION (OR/AND SPOTTING) STAGE (if any): None

3.1.4 Learning strategy
Learning strategy applied FOR GESTURE RECOGNITION (OR/AND SPOTTING) STAGE (if any): Using ConvNets to learn.

3.1.5 Other techniques
Other technique/strategy used not included in previous items FOR GESTURE RECOGNITION (OR/AND SPOTTING) STAGE (if any): We used Improve Depth Motion Maps (IDMM) for the input of ConvNets to learn the features.

3.1.6 Method complexity
Method complexity FOR GESTURE RECOGNITION (OR/AND SPOTTING) STAGE: Real-time

3.2 Data Fusion Strategies
List data fusion strategies (how different feature descriptions are combined) for learning the model / network: Single frame, early, slow, late. (if any): None

3.3 Global Method Description
• Which pre-trained or external methods have been used (for any stage, if any): We use pre-trained models on ILSVRC-2012 for Alexnet.

• Which additional data has been used in addition to the provided ChaLearn training and validation data (at any stage, if any): We only used depth data.
• Qualitative advantages of the proposed solution: simple yet effective
• Results of the comparison to other approaches (if any):

Table 1: Comparative accuracy of proposed method and baseline methods on the ChaLearn LAP ConGD Dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Set</th>
<th>Mean Jaccard Index $J_S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFSK</td>
<td>Validation</td>
<td>0.0918</td>
</tr>
<tr>
<td>MFSK+DeepID</td>
<td>Validation</td>
<td>0.0902</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>Validation</td>
<td><strong>0.2403</strong></td>
</tr>
<tr>
<td>MFSK</td>
<td>Testing</td>
<td>0.1464</td>
</tr>
<tr>
<td>MFSK+DeepID</td>
<td>Testing</td>
<td>0.1435</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>Testing</td>
<td><strong>0.2655</strong></td>
</tr>
</tbody>
</table>

• Novelty degree of the solution and if is has been previously published: incremental

4 Other details

• Language and implementation details (including platform, memory, parallelization requirements): Matlab + Caffe + Python. GPU memory required no less than 3 GB.
• Human effort required for implementation, training and validation? Easy.
• Training/testing expended time? Less than one hour.
• General comments and impressions of the challenge? what do you expect from a new challenge in face and looking at people analysis?: Very good and hope to see more contestants.