Multi-sensor accelerometer-based gesture recognition

Mgr. Matej Králik 1  Mgr. Vladimir Boža PhD. 1
1Department of Computer Science, Faculty of Mathematics, Physics and Informatics, Comenius University

Motivation
Recognizing hand gestures using accelerometers or IMUs has several use cases including: Remote control, health monitoring, gaming, virtual and augmented reality and biometrics. The current state-of-the-art typically uses a single handheld sensor and a vocabulary of very simple gestures. These gestures typically involve a single movement in a specific direction. The use of multiple sensors located on the hand significantly expands the space of recognizable gestures, increases recognition accuracy and allows the use of more natural hand gestures.

Single gestures

Multi gestures

Figure 1: Comparison of single-sensor gestures and multi-sensor gestures.

WAVEGLOVE

In our work we built a hardware glove prototype called WaveGlove. The WaveGlove prototype uses five inertial measurement units (IMUs) attached to a left-hand glove—one on each finger—to record accelerometer and gyroscope data at the rate of 40 Hz.

Figure 2: The WaveGlove prototype.

To the best of our knowledge WaveGlove to be the first device of its kind, which was used to record a large-scale (over 11000 gesture instances) publicly available dataset.

Classification methods

To put our results into context, we study a multitude of datasets and reproduce several previously published methods. We further propose a novel classification method which uses the self-attention mechanism and is based on the Transformer architecture (popularized by Vaswani et al. [1] in 2017 by establishing a new state-of-the-art for machine translation tasks). The Transformer architecture has been previously applied in our field only marginally.

Table 1: Evaluation of classification methods on various datasets with predefined folds. The "*" symbol denotes that the given method was not applied on the dataset. Note that this table only showcases the most relevant methods and datasets—for the full set of results please refer to our work.

<table>
<thead>
<tr>
<th>Method</th>
<th>USC-MHAD</th>
<th>USC-HAD</th>
<th>UTD-MHAD1</th>
<th>UTD-MHAD2</th>
<th>WHARF</th>
<th>WISDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Decision Tree</td>
<td>97.36</td>
<td>80.78</td>
<td>72.74</td>
<td>72.24</td>
<td>79.89</td>
<td>87.09</td>
</tr>
<tr>
<td>Deep convolution LSTM</td>
<td>97.57</td>
<td>80.86</td>
<td>72.79</td>
<td>72.29</td>
<td>79.90</td>
<td>87.10</td>
</tr>
<tr>
<td>DCNN Ensemble (2019)</td>
<td>97.76</td>
<td>81.05</td>
<td>73.28</td>
<td>72.61</td>
<td>79.91</td>
<td>87.10</td>
</tr>
<tr>
<td>Self-attention with sensor embedding (newly proposed)</td>
<td>99.40</td>
<td>99.99</td>
<td>90.95</td>
<td>89.83</td>
<td>76.32</td>
<td>78.69</td>
</tr>
<tr>
<td>Transformer (newly proposed)</td>
<td>94.53</td>
<td>84.69</td>
<td>76.99</td>
<td>72.51</td>
<td>79.92</td>
<td>87.09</td>
</tr>
</tbody>
</table>

Average accuracy

Using the single index finger sensor we observe the best performance, with 99.40% average accuracy.

Figure 4: Confusion matrices generated by using the data only from one of the sensors on the glove.

To support the claim that the recognition accuracy varies depending on the sensor location and gesture types, we show the confusion matrices in Figure 4.

Index finger sensor

Pinky sensor

Comparing the average recognition accuracy when using 1 to 5 sensors of the glove, we find that using up to three sensors significantly improves the accuracy on the Multi-gestures dataset.

Utilizing more than one sensor causes little to no improvement when classifying the Single gestures, where each of the fingers performs the same movement.

Single gestures

Multi gestures

Figure 3: Average accuracy based on the amount of sensors used and the training set size. Single gestures are simple movements, while Multi gestures are more complex and can require a different movement for each finger.

To support the claim that the recognition accuracy varies depending on the sensor location and gesture types, we show the confusion matrices in Figure 4.

• We present a custom hardware prototype, which we used to acquire a dataset of over 11000 gesture instances. To the best of our knowledge this is the largest publicly available dataset of multi-sensor hand gestures. Available at https://github.com/Zajozor/waveglove.

• In an effort to standardize the state-of-the-art, we used 11 human activity recognition datasets, implement several previously published methods and compare their performance.

• We propose a novel Transformer-based network architecture, which shows promising classification results.

• In an ablation study, we show that (only) relevantly designed gestures benefit from the use of multiple sensors. Furthermore, when using only a single sensor the recognition accuracy highly depends on the location of the sensor and gesture type.

Contributions

References


WaveGlove

Figure 3: Average accuracy based on the amount of sensors used and the training set size. Single gestures are simple movements, while Multi gestures are more complex and can require a different movement for each finger.