Multi-sensor accelerometer-based gesture recognition

Mgr. Matej Králik [1] Mgr. Vladimír Boža PhD. [1]

[1]Department 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.

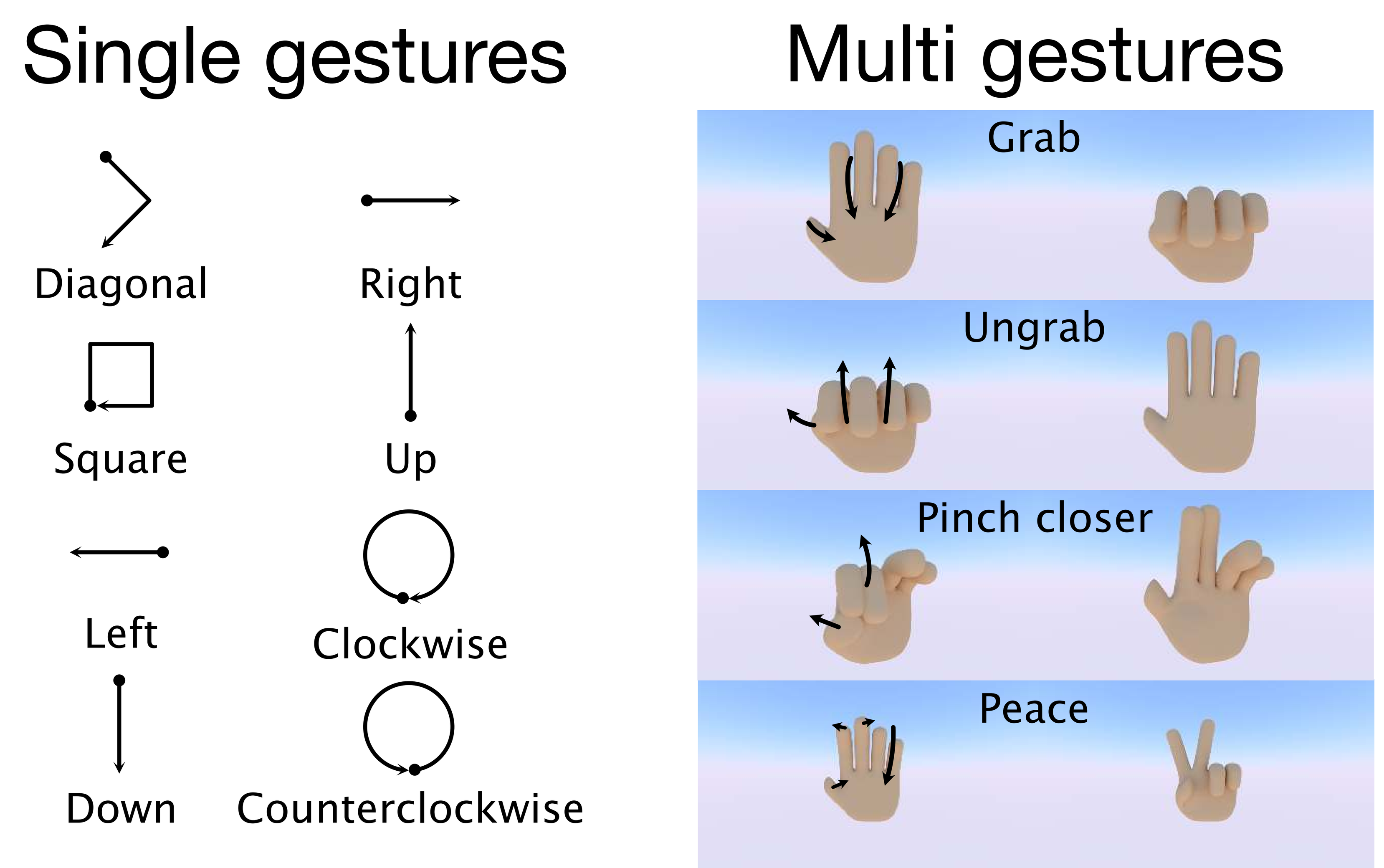

Figure 1: Comparison of single-sensor gestures and multi-sensor gestures.

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.

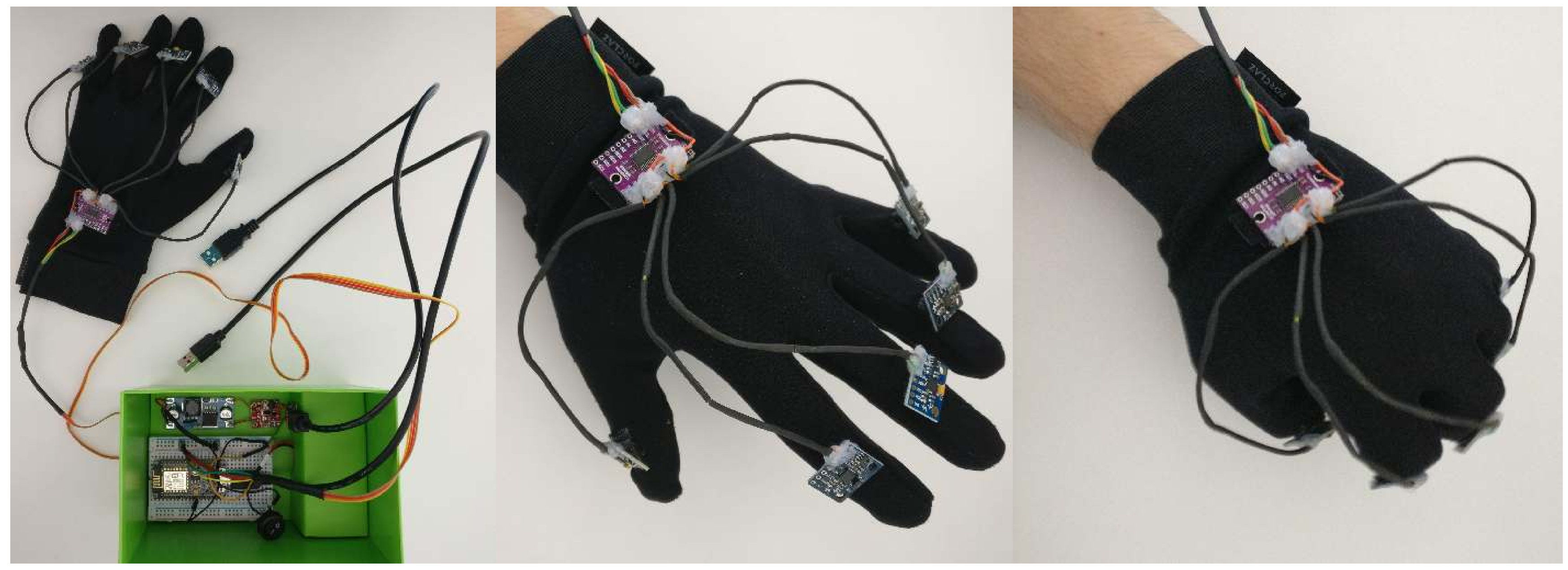

Figure 2: The WaveGlove prototype.

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.

Method	Accuracy (%)								Avg.
	WaveGlove-single	WaveGlove-multi	MHEALTH	USC-HAD	UTD-MHAD1	UTD-MHAD2	WHARF	WISDM	
Baseline Decision Tree	99.10	96.63	**93.41**	89.22	67.51	81.90	66.26	61.61	81.96
DeepConvLSTM (2016)[2]	98.05	99.30	81.01	83.97	67.29	86.50	67.98	**91.23**	84.41
DCNN Ensemble (2019)[3]	-	-	93.09	88.49	62.03	81.63	75.50	89.01	-
Self-attention with sensor embedding (newly proposed)	**99.40**	**99.99**	90.35	**89.83**	**76.32**	**88.42**	**78.63**	84.53	**89.68**

Table 1: Evaluation of classification methods on various datasets with predefined folds. The "-" symbol denotes that the given method was not be 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.

Attribution of multiple sensors

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.

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.

Recognition accuracy using only one of the sensors

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.

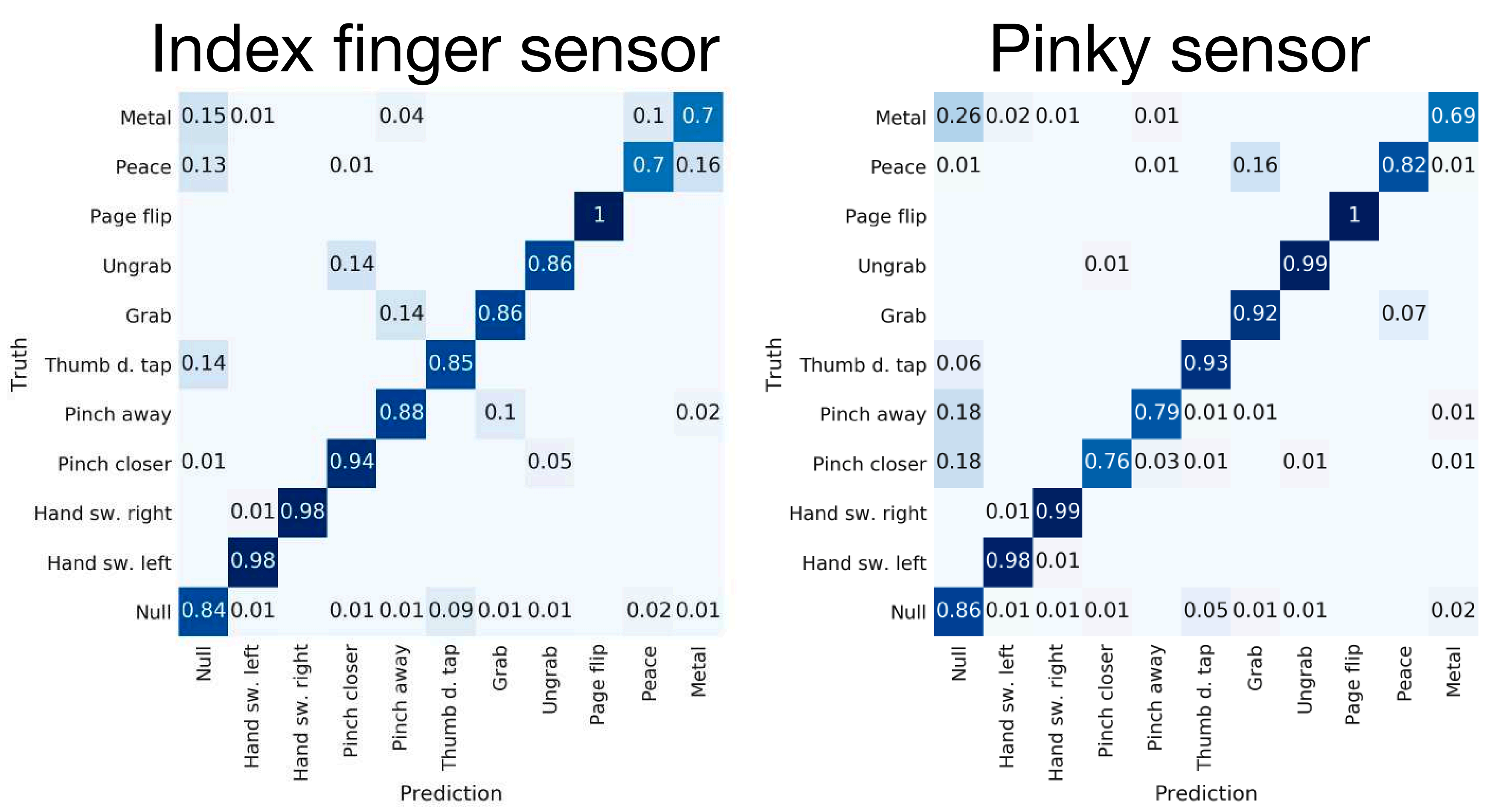

Figure 4: Confusion matrices generated by using the data only from one of the sensors on the glove.

Using the index finger sensor, the *Metal* and *Peace* gestures are classified as the *Null* gesture because they do not include index finger movement. The pairs *Ungrab/Pinch closer* and *Grab/Pinch away* use the same index finger movement, resulting in a recognition confusion.

Using the pinky sensor, the *Peace* and *Grab* gestures are often confused. Several other gestures are confused with the *Null* gesture, because they do not involve significant pinky movement.

Contributions

- 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 use 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.

References

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017.

[2] Francisco Javier Ordóñez and Daniel Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1):115, 2016.

[3] Jessica Sena and William Schwartz. Human activity recognition based on wearable sensors using multiscale dcnn ensemble. In *Anais Estendidos da XXXII Conference on Graphics, Patterns and Images*, page 112–118, Porto Alegre, RS, Brasil, 2019. SBC.