Curation of datasets tailored for HumanE-AI multimodal perception and modelling

Mathias Ciliberto, Prof. Daniel Roggen



Wearable Technologies Lab University of Sussex, Brighton, UK



Our expertise

- Multimodal data collection
- HAR pipeline development
- Wearable sensors







The Problem 1-3 Switch

Door

[3]

[1]

Door





- Scattered datasets
- Difficult to find
- Not always HumanE-AI ready:
 - Little documentation
 - Missing information (e.g. sample)

frequency, formatting)

Different and unclear formats •

[1] Roggen, Daniel, et al. "Collecting complex activity datasets in highly rich networked sensor environments." 2010 Seventh international conference on networked sensing systems (INSS). IEEE, 2010. [2] Zappi, Piero, et al. "Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection." European Conference on Wireless Sensor Networks. Springer, Berlin, Heidelberg, 2008.

[3] Cuspinera, L. Ponce, et al. "Beach volleyball serve type recognition." Proceedings of the 2016 ACM International Symposium on Wearable Computers. 2016.

The Goal



- Advancing HumanE-Al research
- Increase visibility
- Improve quality of datasets

The Solution

- Collection of curated datasets
- Uploaded on AI4EU platform
- For researchers, made by researchers
- Well documented
- Easy-to-use formats
- Integration with AI4EU Experiments









Our datasets - OPPORTUNITY



- Activities of daily living
- 12 subjects
- >30000 interactions
- 17 gestures
- Sensor rich:
 - 24 on-body sensors
 - 12 object sensors
 - 21 ambient sensors

Our datasets – Skoda

- Car manufacturing gestures
- 7 IMU Sensors
- 20 activities
- 2 subjects
- 560 minutes of data



Stiefmeier, Thomas, et al. "Wearable activity tracking in car manufacturing." IEEE Pervasive Computing 7.2 (2008): 42-50.

Our datasets – SHL





- Locomotion and transportation dataset
- 4 Smartphones
 - 15 sensors
- 1 cameras
- 3 Users
- 750 Hours of labelled data (per phone)
- 950 GB of raw data from sensors
- 86075 pictures

Our datasets – Beach Volleyball

- Sport skill assessment
- Collaborative
- 4 type of serving + game
- 4 IMU sensors per person
- 12 subjects



Dataset attributes

- Multimodal
 - Smartphones
 - Wearable sensors
 - Cameras
- Multiple position
 - Different on-body position
 - Environmental sensors
- Multiple people
 - Male, female, different ages, different ethnicity, etc.
- Accurate labels
 - Precise start/end
-Anything else?











Microproject

- Duration: 5-6 PMs
- Tangible outcomes:
 - Curation and publication of datasets on AI4EU web platform
 - Investigate integration with AI4EU Experiments as data source
 - Identify improvements in AI4EU Experiments in order to feed back to developers
 - Publication of Pervasive Computing magazine paper "Best practice and lessons learnt in high value data collection"
- Future development:
 - Development of baseline pipeline in AI4EU Experiments
- Looking for partners with valuable datasets and data collection expertise

Thank you for your attention!

Mathias Ciliberto, m.ciliberto@sussex.ac.uk

Prof. Daniel Roggen daniel.roggen@ieee.org

Curation of datasets tailored for HumanE-AI multimodal perception and modelling