Sound-based transportation mode recognition with smartphones
Lin Wang¹, Daniel Roggen²
¹Centre for Intelligent Sensing, Queen Mary University of London, UK; lin.wang@qmul.ac.uk
²Wearable Technologies Lab, University of Sussex, UK; daniel.roggen@ieee.org

1. Transportation mode recognition

• Applications
  - Context information on user mobility
  - Intelligent service adaptation
  - Individual environmental impact
  - Human-centered activity monitoring

• State-of-the-art [1]
  - GPS
  - Motion (accelerometer, gyroscope, magnetometer)

• Can sound be utilized for transportation mode detection?
  - Microphone in all smartphones
  - Providing rich information on surroundings
  - Already used on sound event recognition
  - Few work on transportation mode recognition
  - Very few transportation datasets contain sound
  - Influence of environmental noise

To answer this question through evaluation on the SHL dataset
  - Machine learning vs deep learning

2. SHL Dataset

• Used in a machine learning challenge in 2018 [3, 4]
• Largest real-world dataset on locomotion/transportation [1,2]
  - 7 months (2812 hours), 17562 km travel distance
  - 3 users & 4 sensor placement
  - 8 transportation modes
  - 16 sensor modalities

• Sound data used in the paper
  - Data of hand-located phone of one user
  - Train: 62 days (271 hours)
  - Test: 20 days (95 hours)

• Exemplary spectrogram
  - (a) without environmental sound
  - (b) with environmental sound

3. Sound-based transportation mode recognition

1. Conv1+Norm1
   ReLu1+Pool1
2. Conv2+Norm2
   ReLu2+Pool2
3. Fullyconnected1
   ReLu3+Drop1
4. Fullyconnected2
   ReLu4+Drop2
5. Input Layer
6. Time domain ZCR (M&V)
7. MFCC1 (M&V)
8. MFCC13 (M&V)
9. Normalize \([0, 1]\)
10. Naïve Bayesian/
    Decision Tree / SVM / Random Forest/ KNN

Sound modalities:
1. Accelerometer
2. Gyroscope
3. Magnetometer
4. Orientation
5. Gravity
6. Linear
7. Pressure
8. Light
9. Battery
10. Sound
11. Wi-Fi
12. GSM
13. GPS
14. Image
15. 10. Sound
16. Google API

Transportation labels:
1. Walk
2. Run
3. Bike
4. Car
5. Bus
6. Train
7. Subway
8. Locomotion
9. Bike
10. Walking
11. Cycling
12. Tram
13. Train
14. Bus
15. Car
16. Walking

Sensor modalities

4. Evaluation results

• Experimental Setup
  - Intel i7-4770@3.4GHz CPU + 32 GB memory
  - GeForce GTX 1080 Ti GPU + 11 GB memory
  - Matlab Machine Learning and Deep Learning Toolbox
  - Training: 52091 frames (5s frames, skip size 20s)
  - Testing: 55818 frames (5s frames, skip size 5s)

• Learning rate
  - 10^-4

• Deep-learning pipeline
  - 5s frames, 64ms sub-frames (half-skip)
  - Conv1+Norm1 & Pool1
  - Conv2+Norm2 & Pool2
  - Fully connected layers
  - ReLu
  - Dropout
  - Classification output

• Classical machine learning pipeline
  - 5s frames, 32ms subframes (half-skip)
  - 28 features per frame: mean and variance of zero-crossing rate and MFCC across subframes

• Sound-based recognition with different classification pipelines
  - CNN performs the best
  - CNN most time consuming
  - Post-processing (PP) can further improve performance

• Sound vs motion [4]
  - Complementary modalities
    - Sound
  - Good at 7 vs 8, 5 vs 6
  - Poor at 1 vs 2 vs 3
  - Motion
    - Good at 1 vs 2 vs 3
    - Poor at 7 vs 8, 5 vs 6

5. Conclusion

• Feasibility of sound-based transportation mode recognition
• Complementary between sound and motion
• Multimodal fusion as future work

References

Acknowledgment
This work was supported by the HUAWEI Technologies within the project “Activity Sensing Technologies for Mobile Users”.

CIS centre for intelligent sensing
University of Sussex

SHL DATASET
Sussex-Huawei Locomotion-Transportation Dataset
http://www.shl-dataset.org