Sound-based transportation mode recognition with smartphones

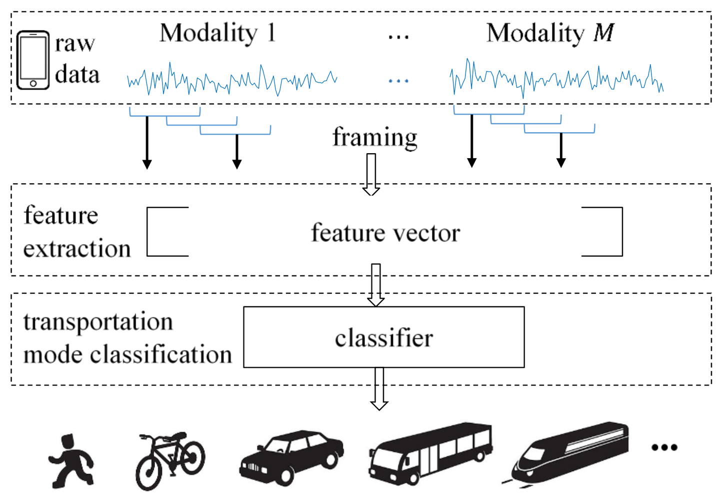
Lin Wang[1], Daniel Roggen[2]

[1]Centre for Intelligent Sensing, Queen Mary University of London, UK; lin.wang@qmul.ac.uk

[2]Wearable 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

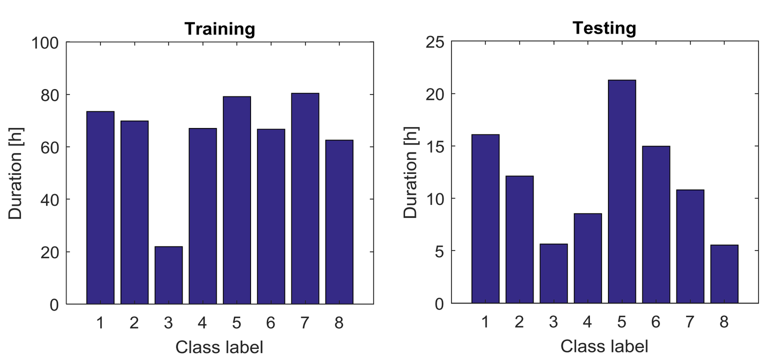
Sensor modalities

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

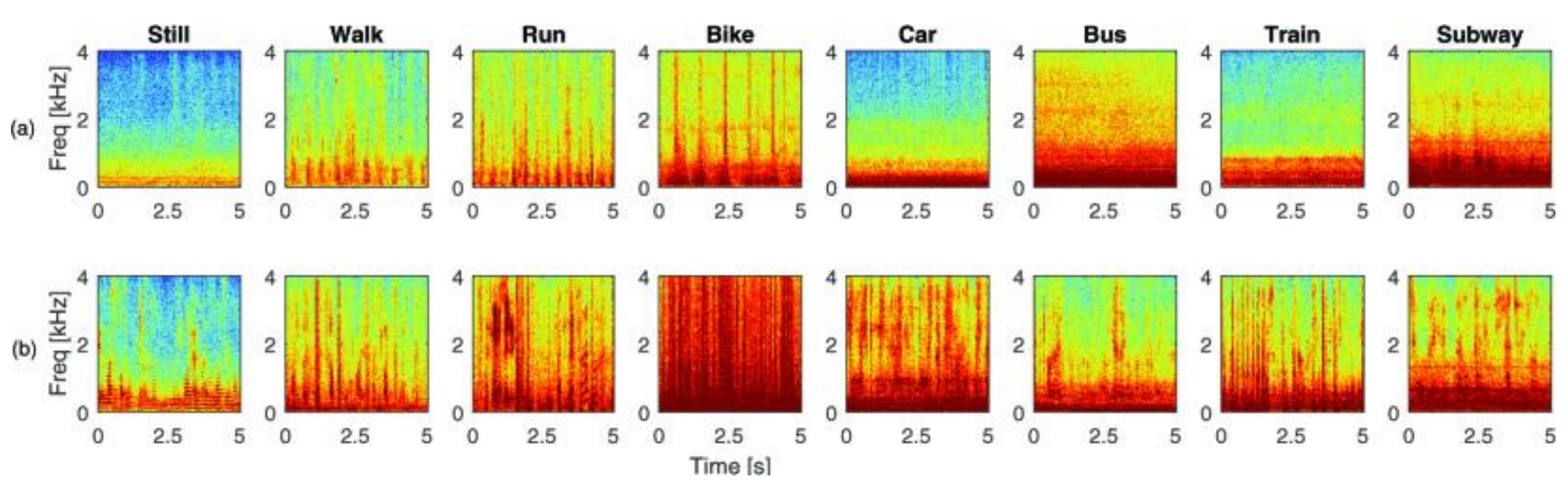
Transportation labels

1. Still	5. Bus
2. Walk	6. Car
3. Run	7. Train
4. Bike	8. Subway

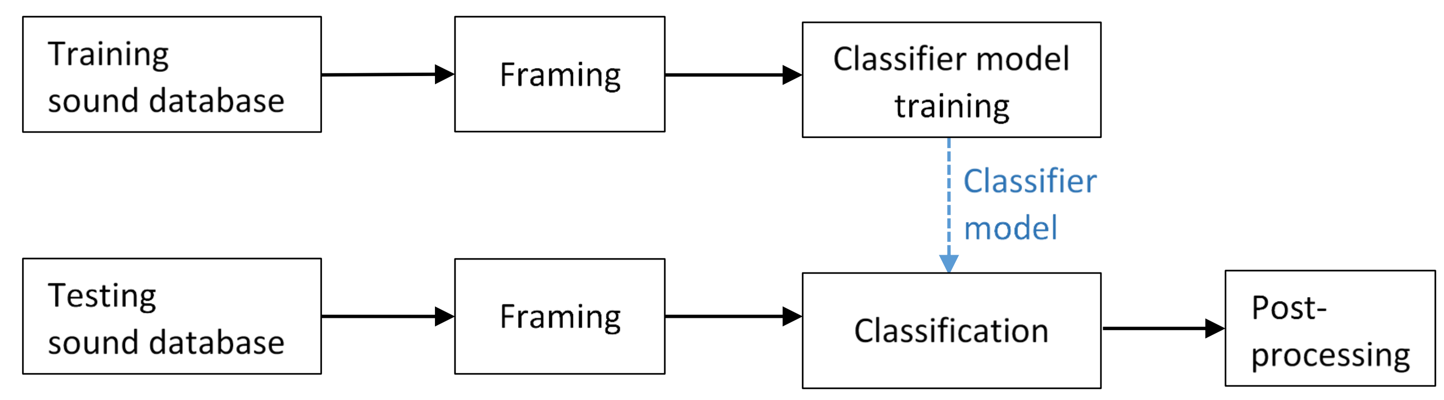
- Sound data used in the paper
  - Data of hand-located phone of one user
  - Train: 62 days (271 hours)
  - Test: 20 days (95 hours)



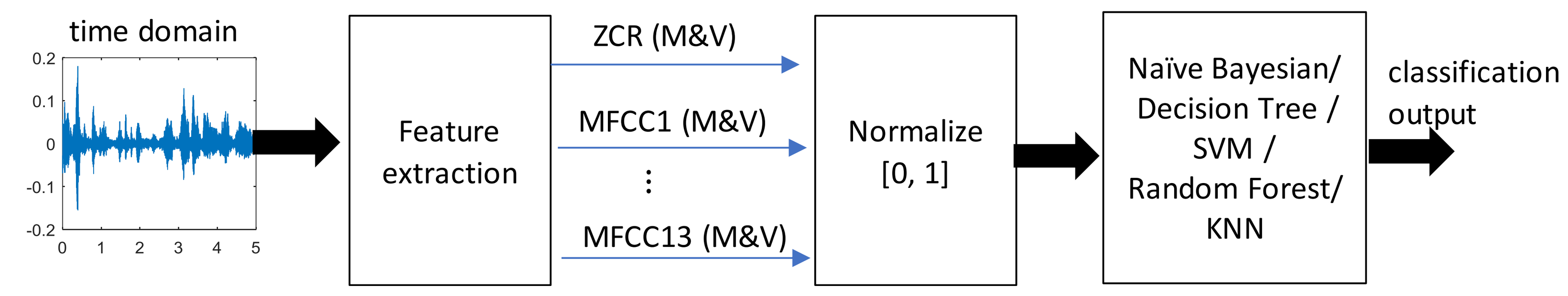
- Exemplary spectrogram
  - (a) without environmental sound
  - (b) with environmental sound



3. Sound-based transportation mode recognition



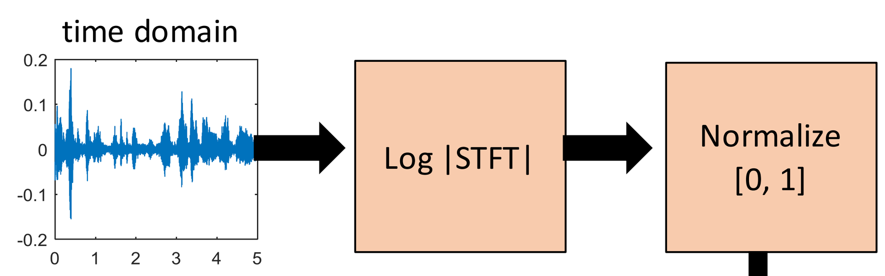
- Classical machine learning pipeline
  - 5s frames, 32ms subframes (half skip)
  - 28 features per frame: mean and variance of zero-crossing rate and MFCC across subframes

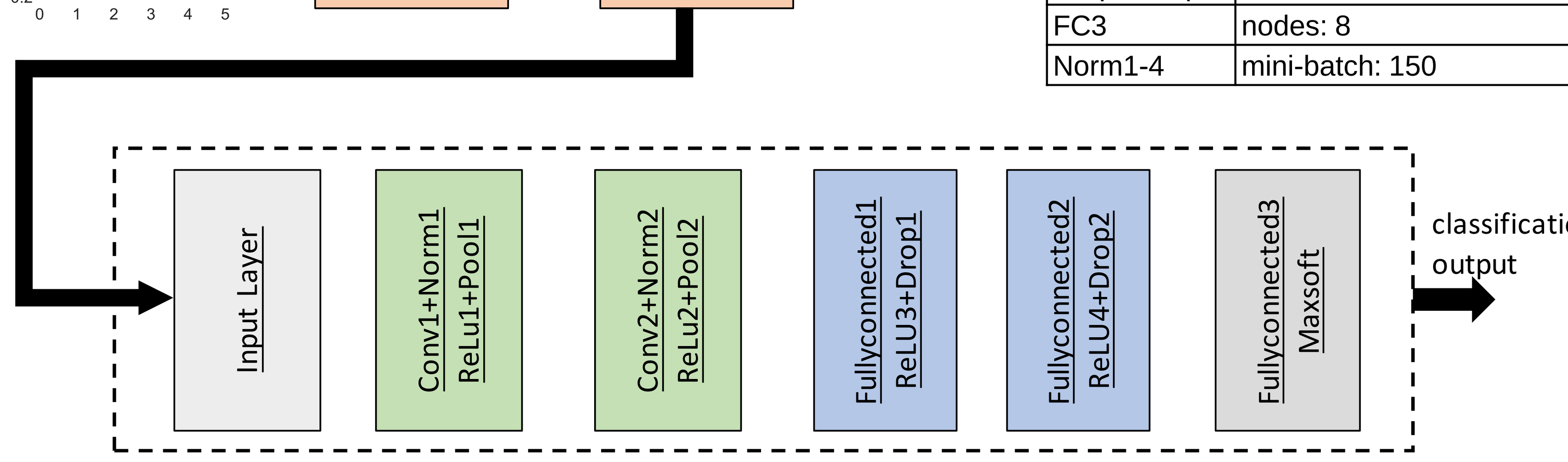


- Deep-learning pipeline
  - 5s frames, 64ms subframes (half skip)

CNN parameters

Input layer	size: (257, 127)
Conv1/Conv2	number: 32; size: (5,5); stride: (1,1); padding: (0,0)
Pool1/Pool2	max pooling: (2,2); stride: (1,1); padding: (0,0)
FC1/FC2	nodes: 300
Drop1/Drop2	50%
FC3	nodes: 8
Norm1-4	mini-batch: 150





- Post-processing
  - To exploit temporal continuity between frames
  - Majority voting across 12 frames per 1 minutes

4. Evaluation results

- Experimental Setup
  - Intel i7-4770@3.4 GHz 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)
  - Evaluation measure: accuracy & F1 score
- Sound-based recognition with different classification pipelines

	Performance [%]		Processing time [s]	
Classifier	Accuracy	F1	Training	Testing
NB	54.6	51.5	0.31	0.1
DT	54.8	50.9	0.93	0.03
RF	62.8	58.3	5.0	0.6
KNN	59.4	57.4	0.04	17.1
SVM	58.8	53.0	223.6	0.18
CNN	80.6	77.9	39448	70.3
CNN+PP	86.6	85.6		

- CNN performs the best
- CNN most time consuming
- Post-processing (PP) can further improve performance

- Sound vs motion [4]

Sound (80.6%)

Groud truth \ Prediction	1	2	3	4	5	6	7	8
1	79	13	0	1	0	1	1	4
2	17	65	2	8	1	1	2	4
3	0	24	59	16	1	0	0	1
4	2	10	4	82	1	2	0	0
5	1	1	0	1	92	1	4	1
6	0	0	0	1	4	88	2	4
7	5	2	0	1	1	1	87	2
8	2	1	0	1	0	2	1	94

Motion (83.3%)

Groud truth \ Prediction	1	2	3	4	5	6	7	8
1	89	3	0	1	1	1	4	2
2	3	94	0	1	0	0	0	1
3	0	1	98	1	0	0	0	0
4	2	4	1	91	0	1	0	0
5	2	0	0	0	75	19	2	1
6	2	1	0	1	6	85	2	2
7	6	1	0	1	2	5	67	18
8	5	1	0	0	1	1	27	64

1-Still; 2-Walk; 3-Run; 4-Bike; 5-Bus; 6-Car; 7-Train; 8-Subway

- 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

[1] L. Wang, H. Gjoreski, M. Ciliberto, S. Mekki, S. Valentin, D. Roggen, “Enabling reproducible research in sensor-based transportation mode recognition with the Sussex-Huawei dataset,” IEEE Access, 2019.

[2] H. Gjoreski, M. Ciliberto, L. Wang, F. J. O. Morales, S. Mekki, S. Valentin, D. Roggen, “The University of Sussex-Huawei locomotion and transportation dataset for multimodal analytics with mobile devices,” IEEE Access, 2018.

[3] L. Wang, H. Gjoreski, K. Murao, T. Okita, D. Roggen, “Summary of the Sussex-Huawei locomotion-transportation recognition challenge”, Ubicomp 2018.

[4] L. Wang, H. Gjoreski, M. Ciliberto, S. Mekki, S. Valentin, D. Roggen, “Benchmarking the SHL recognition challenge with classical and deep-learning pipelines”, Ubicomp 2018.

Acknowledgment

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

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