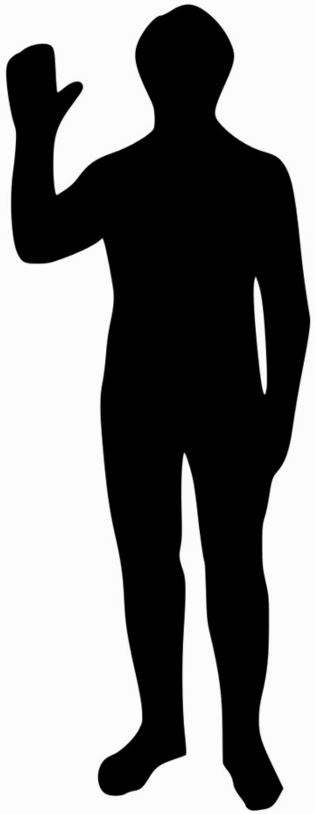
PERVASIVE DATA SCIENCE
SENSING FOR SCIENCE

Petteri Nurmi
petteri.nurmi@helsinki.fi

SENSING

Sight

Smell



Hearing

Taste

Touch

process of perceiving/becoming aware of external or internal stimuli

SENSING AND DEVICES

Inertial sensors: accelerometer and gyroscope

Location tracking: GPS, WiFi

Microphone / voice recognition

Microphone / voice recognition, proximity detection

Heart rate, skin conductivity, other physiological parameters

Air quality sensors, thermal cameras

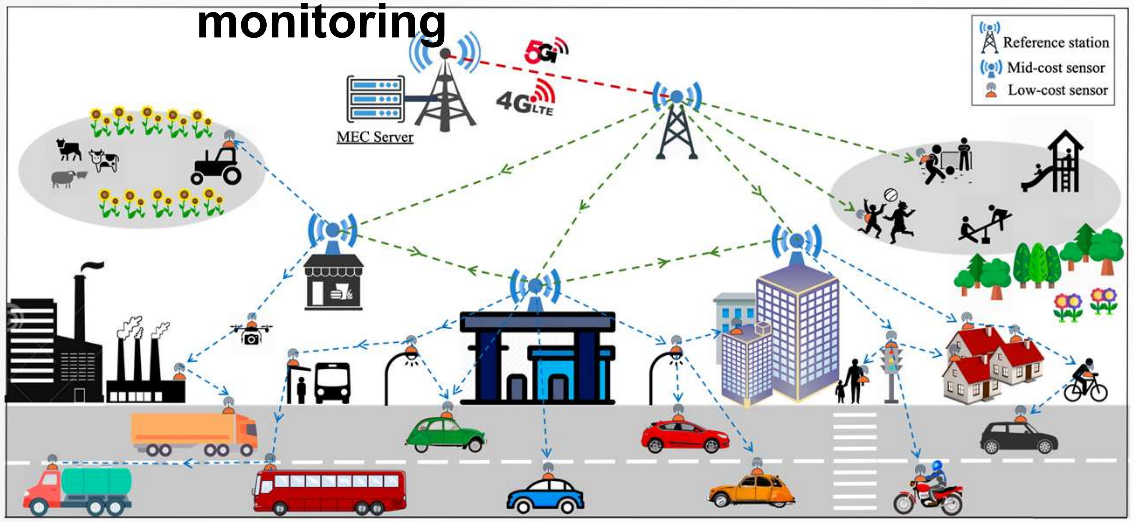
PERVASIVE DATA SCIENCE

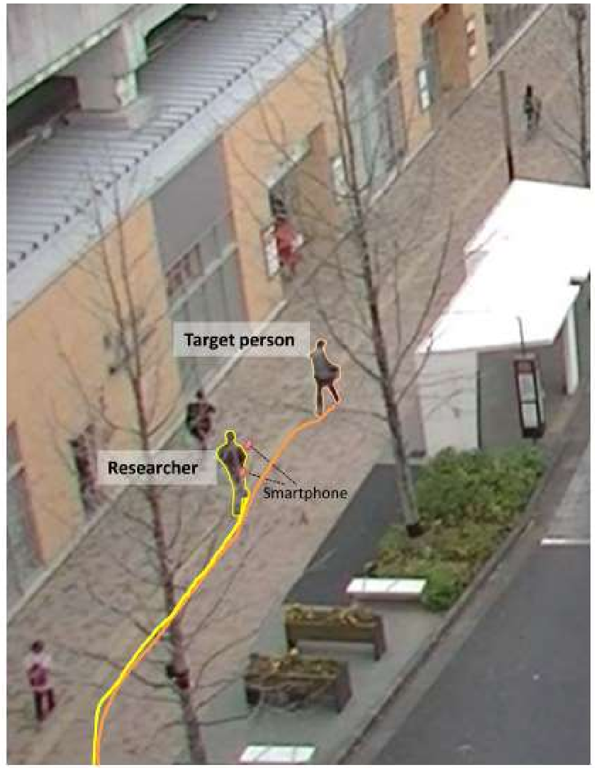
Sensor-enabled devices offer *new way of collecting* information

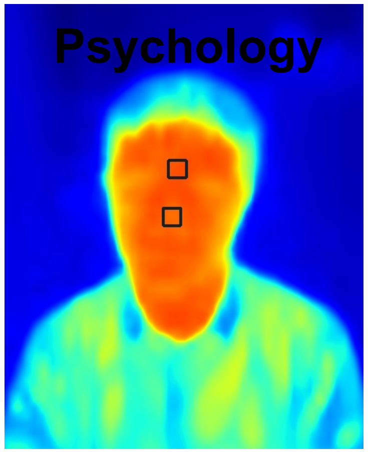
- Devices often carried around everyday situations ➔ powerful mechanism for observing humans and their behaviour
- Potential for *scaling up studies:*
  - Population-scale: reaching larger number of users through the use of off-the-shelf devices
  - Temporal-scale: data collection tools used as part of other activity, potential for capturing longitudinal data
  - Measurement-scale: enable capturing more detailed information
- Pervasive data science = data science using data collected from sensor-enabled devices

SENSING IN SCIENCES

Atmospheric sciences: air quality monitoring







Urban planning and architecture: public space usage sensing

SENSING IN SCIENCES

- Studies take place in everyday life ➔ complex "ecosystem" for data collection
  - Should capture large-amounts of users ➔ need to rely on commodity devices with possibly lower sensor quality
  - Studies should run long periods of time ➔ need to ensure data collection does not interfere with other functionality of devices
- Sensors more powerful than originally imagined
  - Accelerometers originally used to detect orientation of mobile phone, now used for recognizing activities, transportation behaviour, etc.
  - Possibilities to extract new types of information and to run new types of studies!
  - Naturally also raises privacy challenges!

CHALLENGES: VELOCITY AND VOLUME

- Pervasive deployments can generate unprecedented amounts of data
  - Autonomous cars estimated to generate 1GB/s for generating driving decisions
  - Visual processing applications, such as augmented cognition or AR other examples of domains with high volume
- How to support efficiently processing such data volumes?
  - Local processing not sufficiently powerful
  - Volume and velocity of data streams make offloading infeasible

UBISPARK: COMPUTING AS A SERVICE

- Opportunistically takes advantage of *unused computing resources* of (smart) devices within each others proximity
  - Serves as *opportunistic cloudlet* where computing tasks can be provisioned

Lagerspetz et al., IEEE Pervasive, 2019

UBISPARK: EVALUATION

- As case study we have run experiments using distributed image recognition

- MobileNet model running object recognition from continuous video
- Experiments using different combinations of smartphones and a smart TV

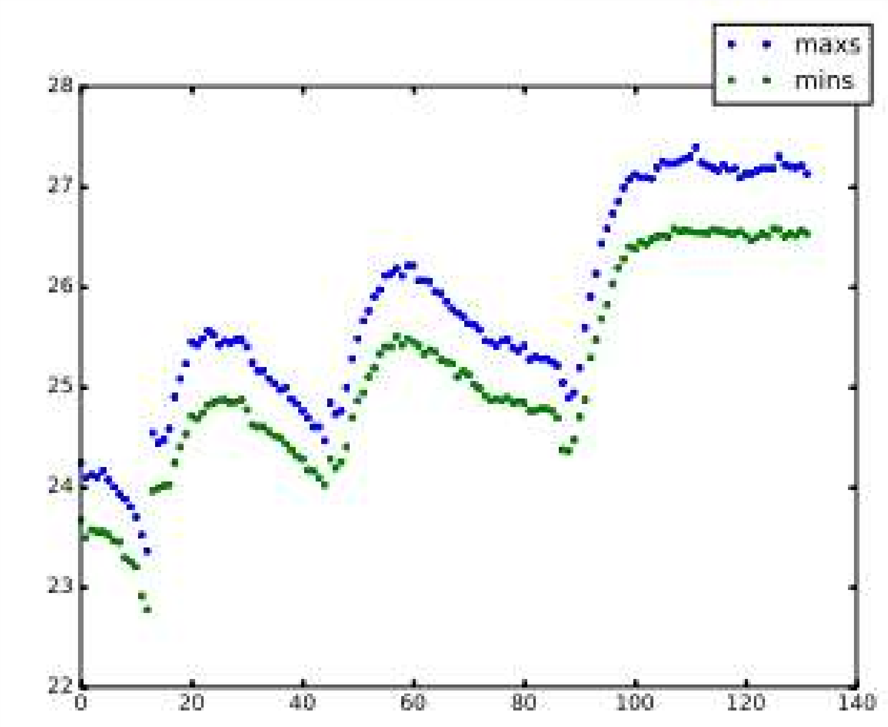
RESULTS

# Devices	Device Model(s)	Time (ms)	Speed (FPS)
1	Philips Smart TV (TV)	702,999	14
1	Samsung Galaxy S4 (S4)	491,018	20
1	OnePlus 3 (OP3)	416,327	24
1	Huawei Honor 7 (Honor7)	266,939	37
2	Honor7 + TV	291,159	34
2	OP3 + TV	243,926	41
2	S4 + TV	226,668	44
2	Honor7 + S4	226,262	44
2	OP3 + S4	201,348	50
2	OP3 + Honor7	182,736	54
3	Honor7 + S4 + OP3	146,658	68
4	Honor 7 + S4 + OP3 + TV	135,644	74
Amazon EC2: t3.medium VM		397,563	25
EC2: c5.2xlarge (CPU only)		197,365	51
EC2: p3.2xlarge (without GPU)		287,364	35
EC2: p3.8xlarge (without GPU)		265,634	38
EC2: p3.2xlarge (with GPU)		132,548	75
EC2: p3.8xlarge (with GPU)		111,329	90

- Using more than one device always improves performance
- With slower devices combined speed close to speed of individual devices
- Compared to Amazon VMs, performance comparable but significantly cheaper!
- Running weaker GPU-enabled devices for a week would cost roughly the same as buying a smart device!

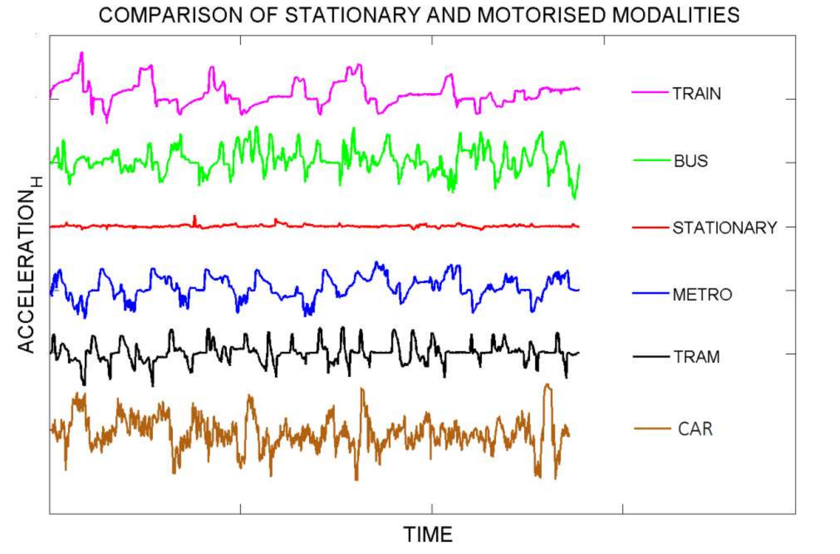
CHALLENGES: SENSOR ACCURACY

Example:
Thermal camera measurements against a wall in a room with stable temperature



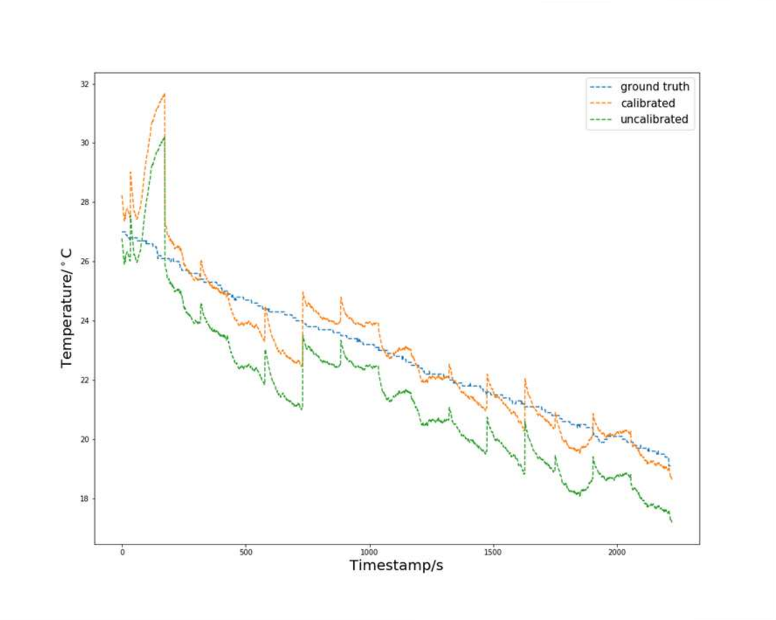
EXAMPLE: SENSOR ACCURACY

Improve Analysis Pipeline



Hemminki et al., ACM SenSys, 2013

Use AI

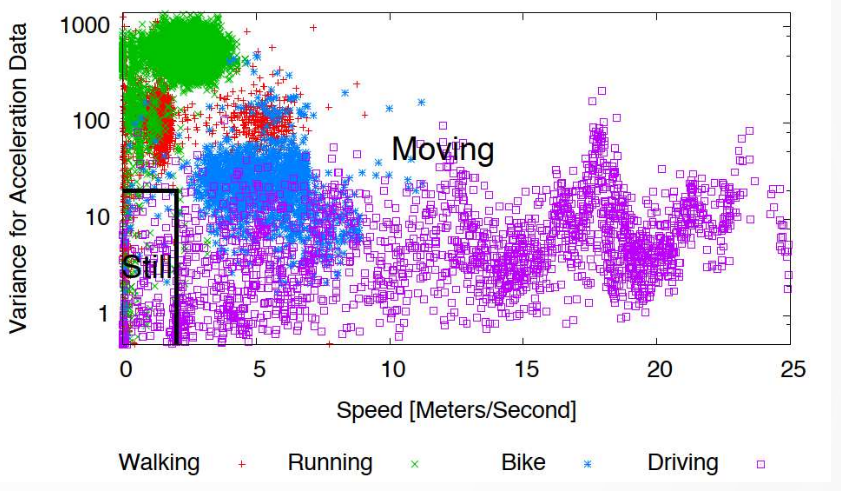


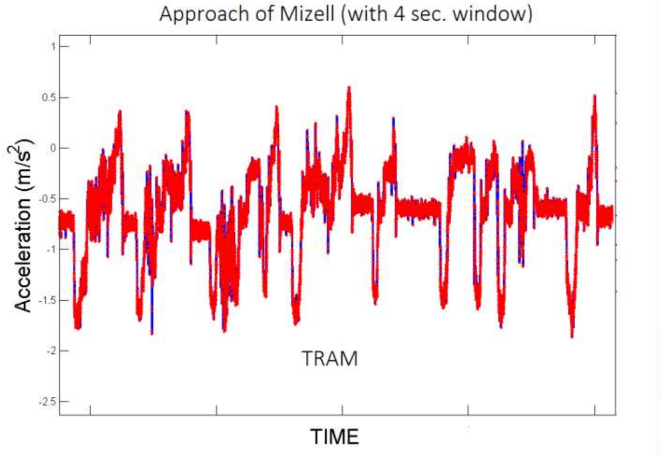
Malmivirta et al., IEEE PerCom, 2019

CHALLENGES: RESOURCE-ACCURACY TRADE-OFF

Example:
Variance-based motion thresholding for removing stationary periods

Bhattacharya et al., 2014

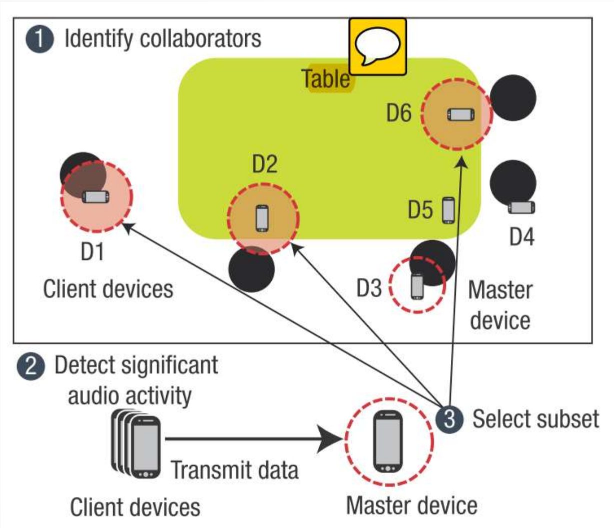


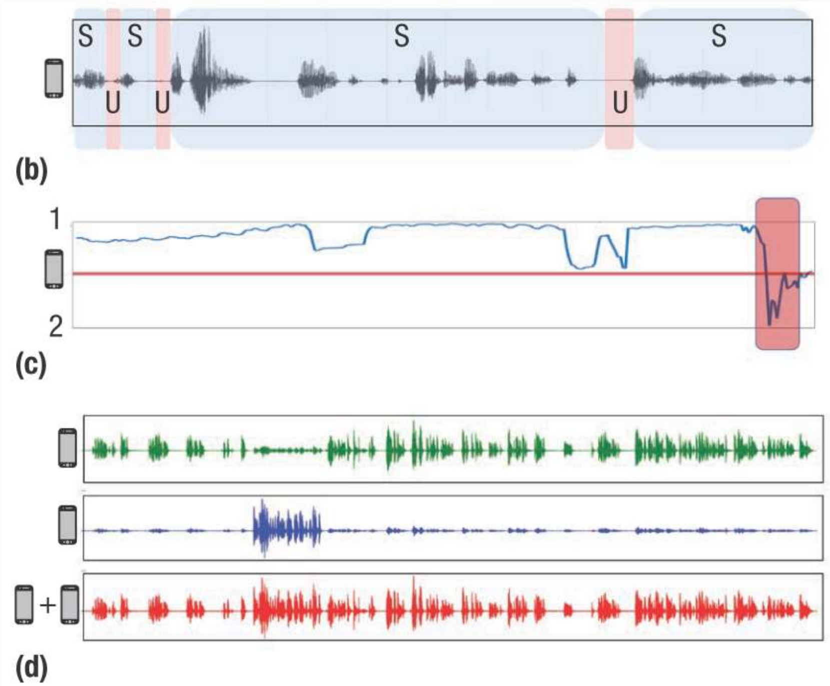


Example:
Low-pass filter gravity elimination from motion measurements

Hemminki et al., 2014

EXAMPLE: RESOURCE-ACCURACY TRADE-OFF

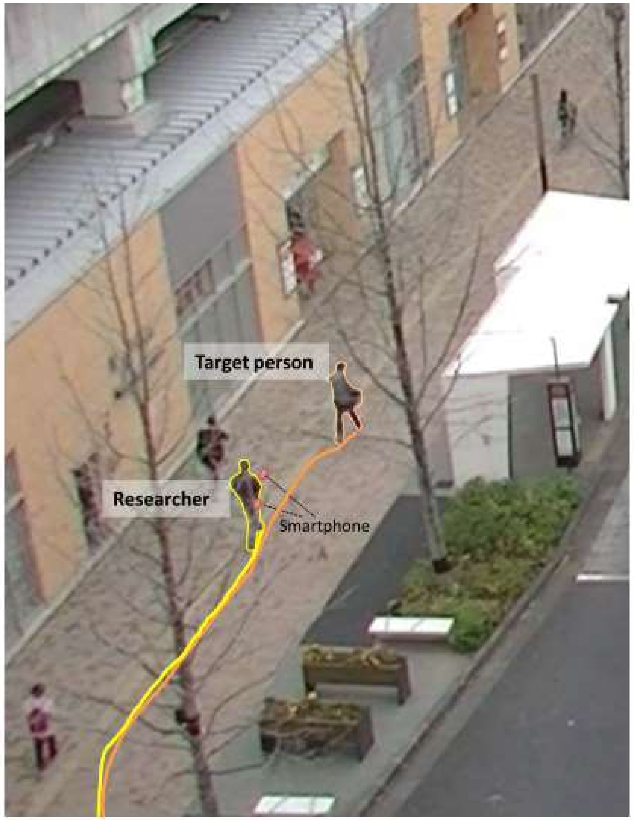




Leppänen et al., IEEE Computer, 2016

PERVASIVE DATA SCIENCE CASE STUDY: CROWD REPLICATION

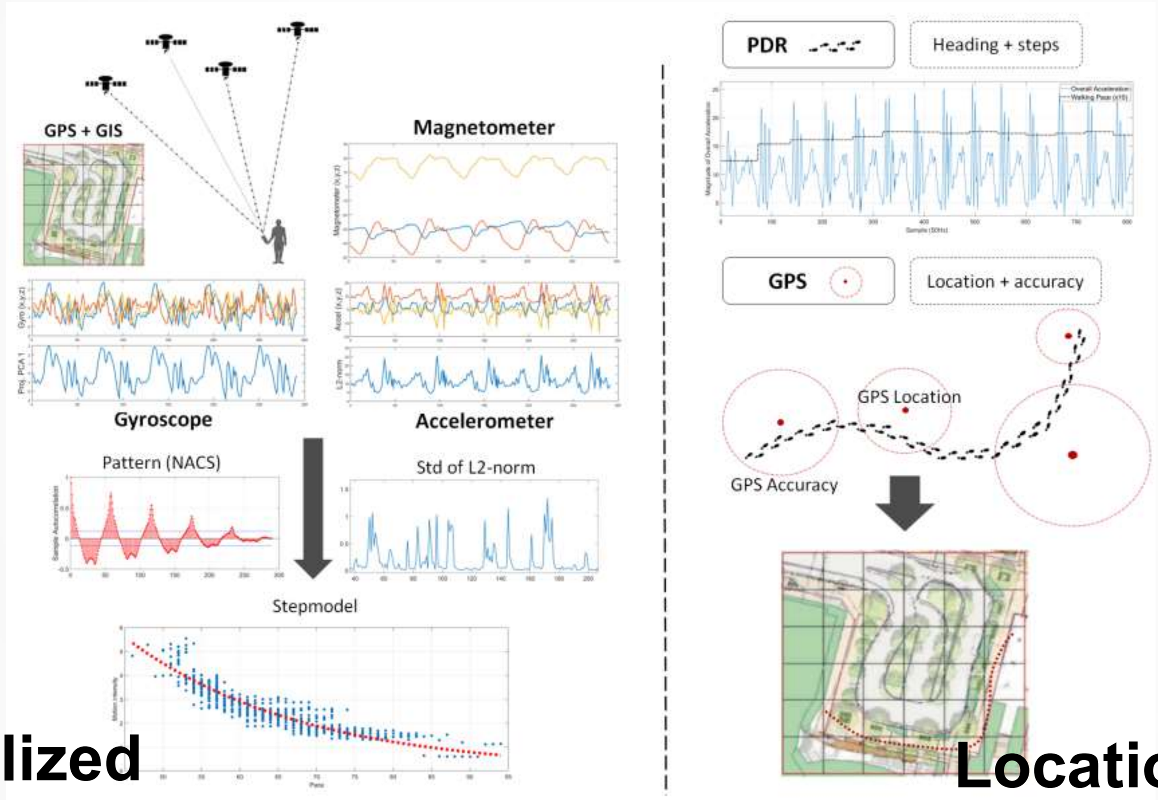
- Method designed to facilitate evaluation of public spaces
- Combines *direct observation* with *sensing*
  - Researchers follow people in an urban space while wearing sensors (or smartphone)
  - Captures fine-grained sensor trace of people together with observations
- Aims at increasing sample sizes and providing finer data granularity than current data collection methods



Hemminki et al., ACM TSAS, 2019

Hemminki et al., SIGSPATIAL, 2016

MOBILE SENSING IN CROWD REPLICATION

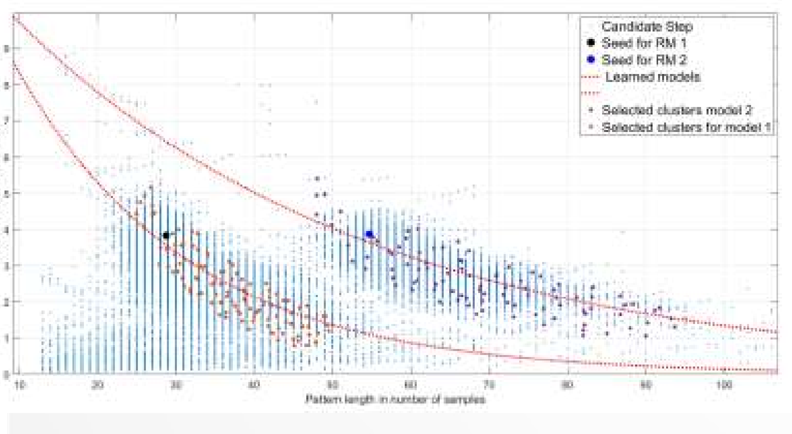


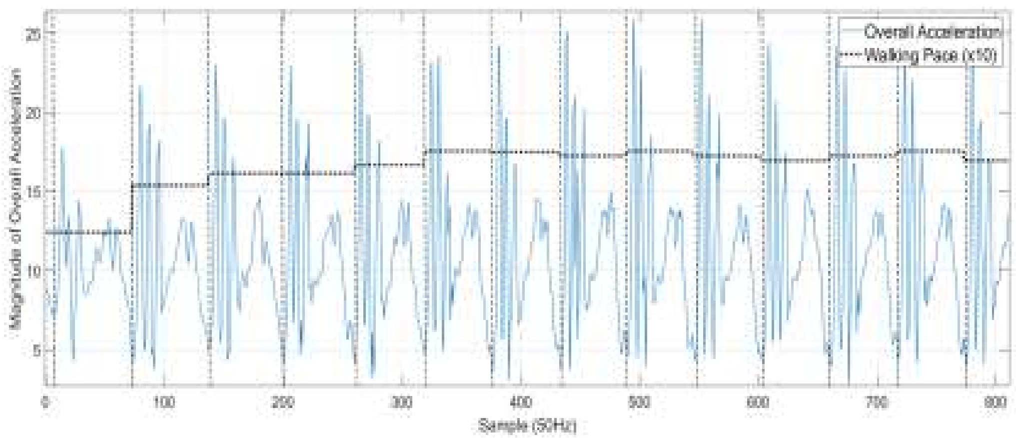
Personalized step modeling

Location Trajectory Estimation

MOBILE SENSING IN CROWD REPLICATION

- Crowd replication relies on *high accuracy* mobile sensing
  - Sensor traces collected from researcher and used to learn a personalized stride model that captures his/her walking characteristics
  - Enables capture of fine-grained motion information and linking it with indicators of human behaviour (extracted through direct observation)
- Accuracy of sensing critical for ensuring validity of observations!
- Sampling theories used to ensure data collection scientifically valid



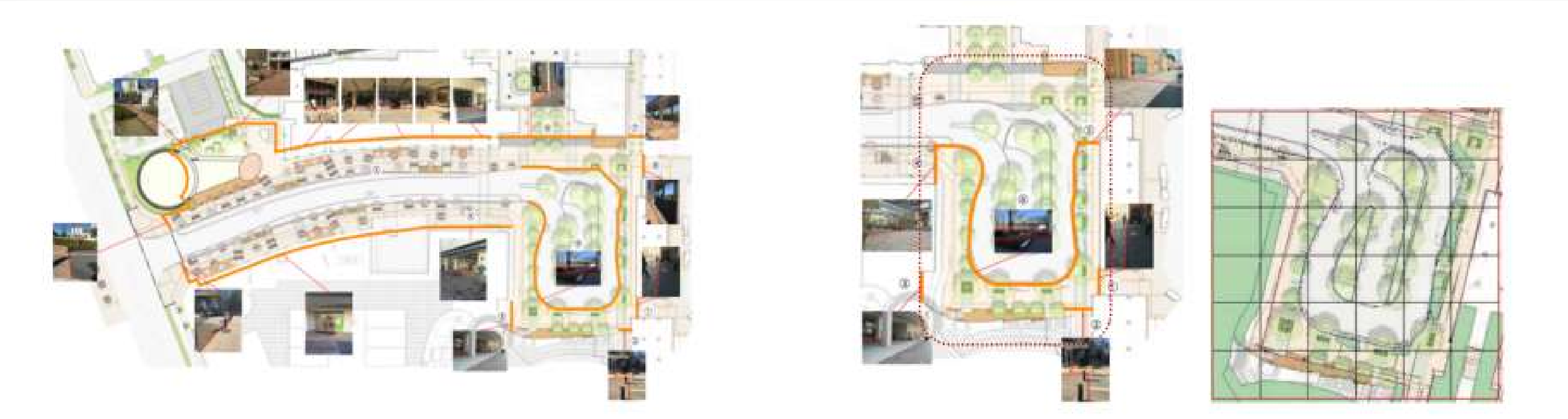


CROWD REPLICATION VS OTHER DATA COLLECTION TECHNIQUES

	Work Effort			Sample	Data resolution	Bias	Privacy risks	Data richness	Recommended for
	Preparation	Data Collection	Analysis						
Direct Observation	small	moderate	large	small	fine grained	small	moderate	rich	Qualitative research (w/ demographic & activity data)
Visual surveillance (manual coding)	large	moderate	very large	large portion in small area	fine grained	small	high	poor	Quantifying short trips in small spaces (exploratory analysis)
Visual surveillance (computer vision)	large	moderate	little	large portion in small area	fine grained	small	high	poor	Quantifying short trips in small spaces
Voluntary location tracking (GPS, GSM or CDR)	large	little	little	small portion in large area	less coarse	mobile users	small	poor	Quantifying city-wide journeys
Location surveillance (GSM or CDR)	small	little	little	good portion in large area	coarse	mobile users	high	poor	Quantifying city-wide journeys
Crowd sensing (wifi, bluetooth, etc.)	moderate	small	little	large portion in medium area	less coarse	mobile users	high	poor	Quantifying trips in large indoor spaces
Crowd replication	small	small	little	large portion in small or mediam area	fine grained	small	moderate	rich	Quantifying trips in neighborhood-scale spaces encompassing multiple streets (w/ demographic & activity data)

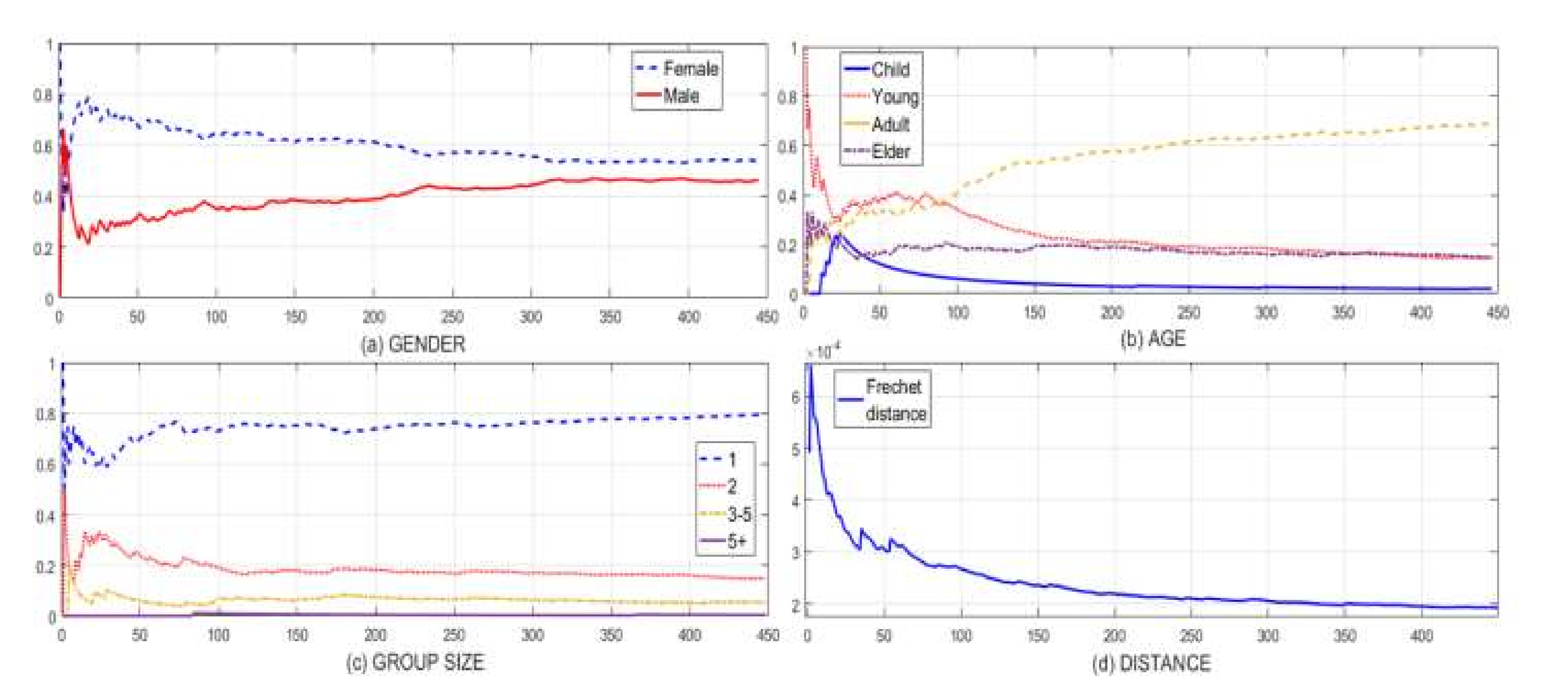
Preserves benefits of direct observation without its main disadvantages, while offering similar advantages as visual surveillance!

EXPERIMENTAL SETUP



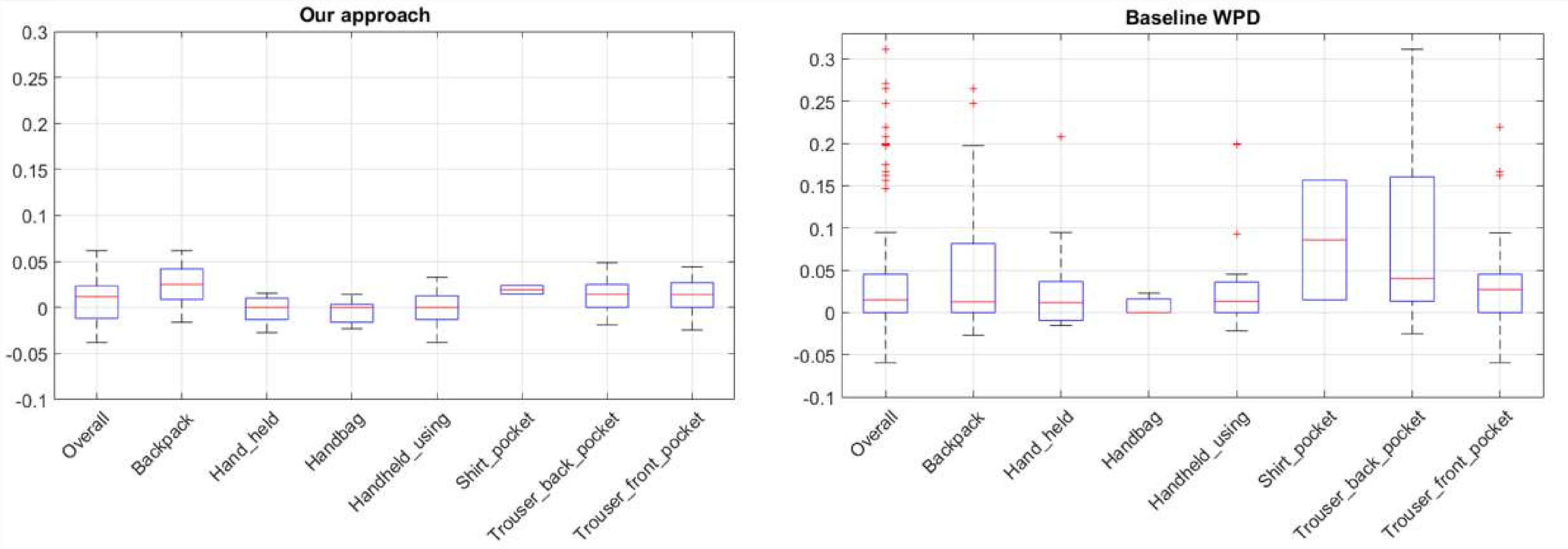
- Target Area: Kashiwano-ha Station Square within Metropolitan Tokyo
- Three trials, one 30 min trial during the entire area and two three hour trials (weekday and weekend) on a subset of the entire space

VALIDITY OF CROWD REPLICATION



- Distribution of key demographics and trajectories converges rapidly to target distribution
- Around 250 replicated users needed for our space

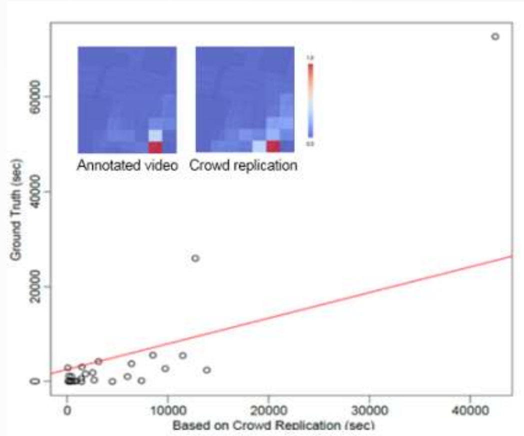
SENSING ACCURACY



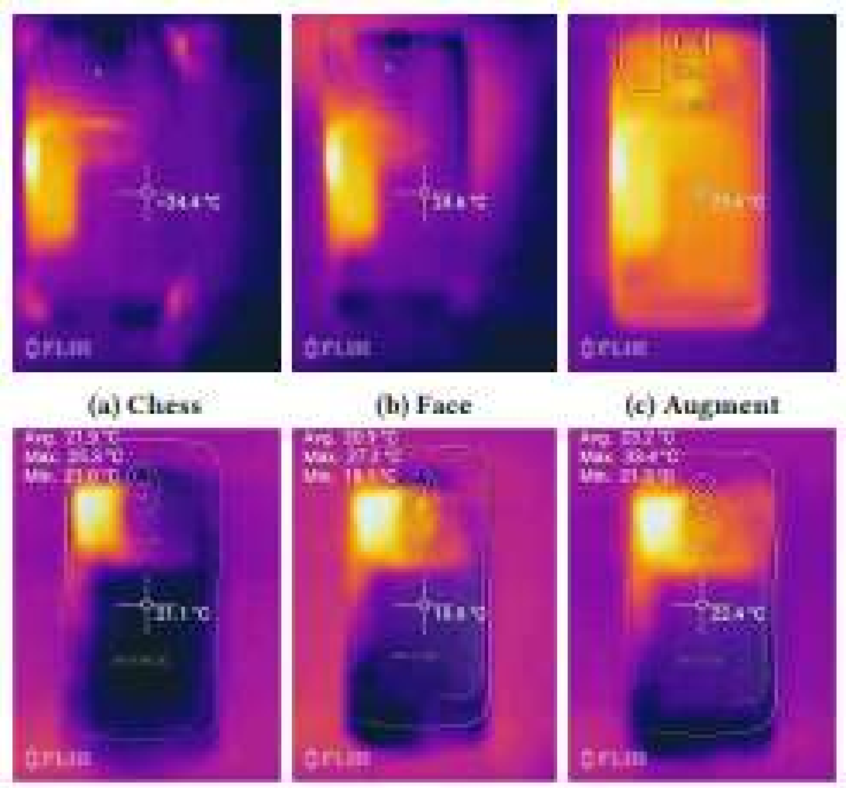
- Median error rate 1.7% (baseline 2.3%)
- Highest errors for upper body placement, errors for positions used in replication small

ACCURACY OF INDICATORS

- Comparison against visual surveillance through *dwell* time
  - I.e., how long people stay within one location
  - 0.914 correlation for dwell index
  - Most errors within one grid cell close to the entry point
  - Mostly due to positioning errors and inaccuracies in visual surveillance
- As another example, we compare stride length between weekdays and weekends
  - Slower stride length during weekends



OTHER EXAMPLES: ENERGY-ESTIMATION THROUGH THERMAL IMAGING



- Estimating device's *relative* energy consumption through thermal camera
- Captures relative differences in energy consumption, even if absolute estimates inaccurate
- Designed for devices with no detachable power source

Flores et al., HotMobile, 2019

OTHER EXAMPLES: AIR QUALITY SENSING

- Low-cost sensor for air quality monitoring
- Advanced AI techniques used to improve accuracy
- 50-90% improvements in highly polluted areas, 22-50% in low pollution areas
- Example of how pervasive data science research can help ***scale up*** sensor deployments

TAKEAWAYS

- Sensors emerging as powerful mechanisms for collecting data
  - Possibility to scale up scientific investigations
  - Possibility for new types of scientific investigations
- Pervasive Data Science refers to use of sensors for scientific investigations
  - Accuracy and long-term operation critical design considerations
- Single sensor can be more powerful than you ever imagine!
- Can support practically **any** scientific domain

REFERENCES

- S. Hemminki, P. Nurmi, S. Tarkoma, Gravity and Linear Acceleration Estimation on Mobile Devices, Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (Mobiquitous), 2014.
- E. Lagerspetz, J. Hamberg, X. Li, H. Flores, P. Nurmi, N. Davies, S. Helal, ”Pervasive Data Science on the Edge”, IEEE Pervasive Computing, 2019
- H. Flores, J. Hamberg, X. Li, T. Malmivirta, A. Zuniga, E. Lagerspetz, P. Nurmi, Evaluating Energy-Efficiency using Thermal Imaging, Proceedings of the 20th International Workshop on Mobile Computing Systems and Applications (HotMobile), 2019
- T. Malmivirta, J. Hamberg, E. Lagerspetz, X. Li, E. Peltonen, H. Flores, P. Nurmi, “Hot or Not? Robust and Accurate Continuous Thermal Imaging on FLIR cameras”, Proceedings of the 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom), 2019.
- S. Hemminki, K. Kuribayashi, S. Konomi, P. Nurmi, S. Tarkoma, ”Crowd Replication: Sensing-Assisted Quantification of Human Behaviour in Public Spaces”, ACM Transactions on Spatial Algorithms and Systems, 2019.
- J. Leppänen, A. Peltonen, H. Guo, S. Hemminki, P. Nurmi, S. Tarkoma, ” Collaborative and Energy-Efficient Speech Monitoring on Smart Devices”, IEEE Computer, 2016.
- S. Bhattacharya, H. Blunck, M. Kjaergaard, P. Nurmi, Robust and Energy-Efficient Trajectory Tracking for Mobile Devices, IEEE Transactions on Mobile Computing (TMC), 2015.
- S.Hemminki, K. Kuribayashi, S. Konomi, P. Nurmi, S. Tarkoma, ” Quantitative Evaluation of Public Spaces using Crowd Replication”, Proceedings of 24th International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL), 2016.
- S. Hemminki, P. Nurmi, S. Tarkoma, ”Accelerometer-Based Transportation Mode Detection on Smartphones”, Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (SenSys, Rome, November 2013).

THANK YOU!

Petteri Nurmi
Associate Professor
Distributed Systems and Internet of Things
Pervasive Data Science: pds.cs.helsinki.fi
petteri.nurmi@helsinki.fi