Human Activity Recognition

이영기
서울대학교 컴퓨터공학부

서울대학교
SEOUL NATIONAL UNIVERSITY

Overview

- **Overview of Activity Recognition**
- Signal Processing
  - Fourier Transformation (FT)
  - Filters (Noise Removal)
  - Feature Extraction
- State-of-the-art Techniques
  - IMU sensor based
    - Gesture Recognition
    - Activity Recognition
  - WiFi signal based Activity Recognition
  - Vision based Activity Recognition
  - Audio based Activity Recognition

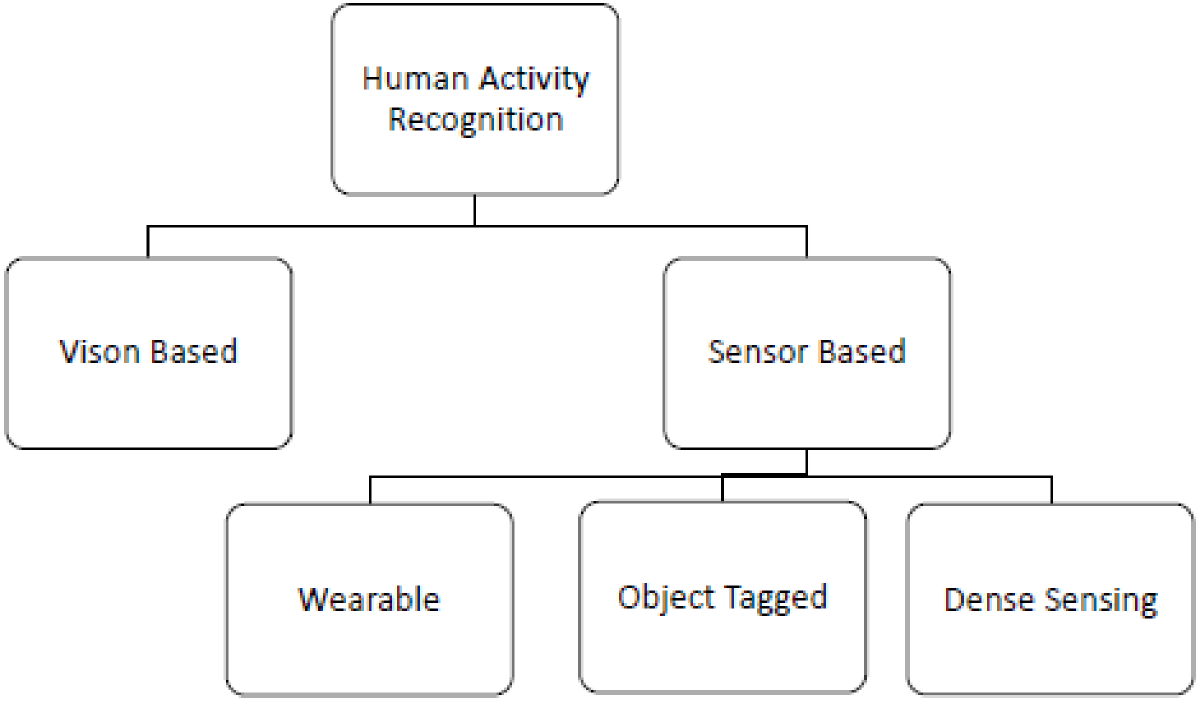
Mobile Activity Tracker

- Everyday exercise progress monitor and motivator
- Provide reliable feedback about how much they move (People often overestimate!)
- Provide instant and constant feedback about activity levels
- Gamify to encourage individuals to compete in getting fit and losing weight

Human Activity Recognition (HAR)

- Identification of the specification movement or action of a person based on a sensor data.

Approaches for HAR



Sensor Types for Activity Recognition

(a) Surveillance Cameras
(b) Depth Cameras
(c) Wi-Fi
(d) Accelerometer
(e) Gyroscope
(f) Proximity Sensor
(g) RFID

(a)

(b)

(c)

(d)

(e)

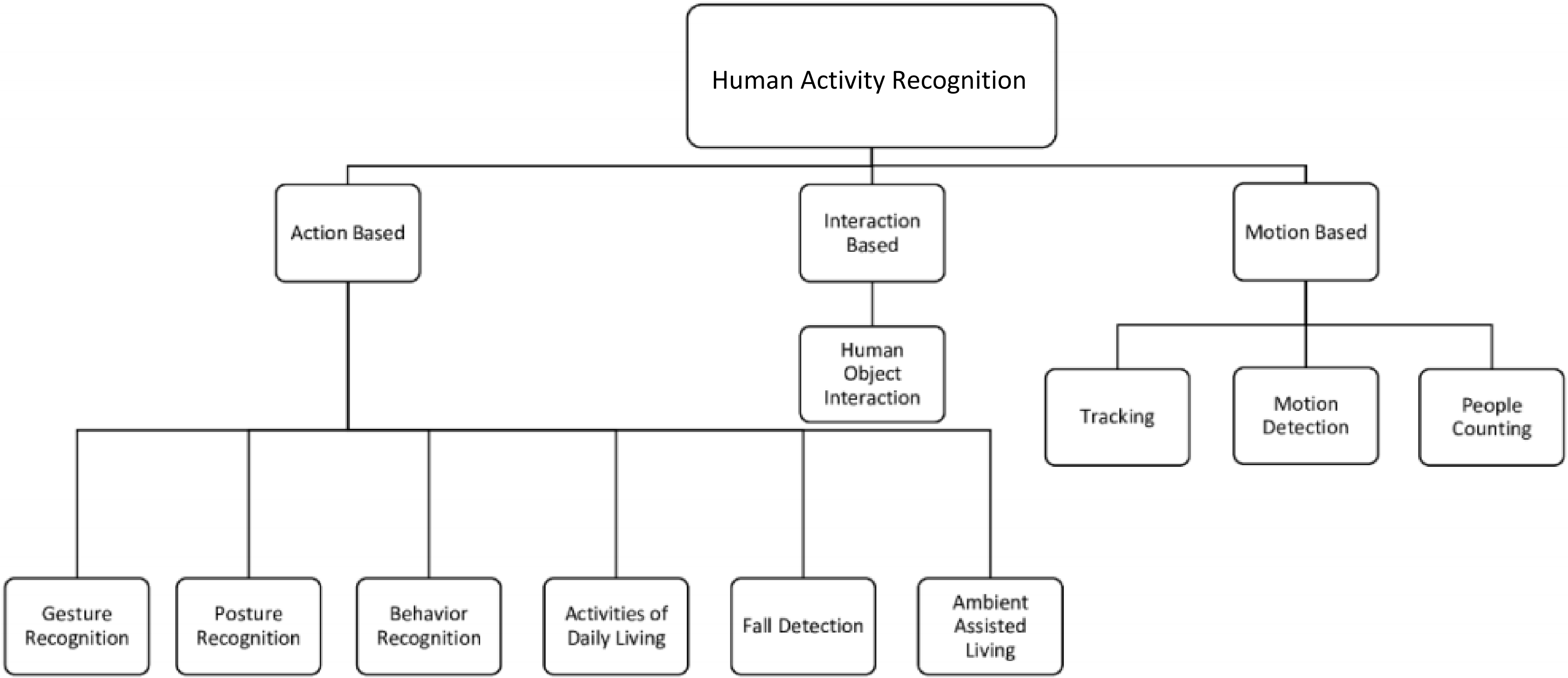
(f)

(g)

Things to Consider

- Approach
- Technology
- Information Type
- Machine Learning Algorithm Used
- Supervised/Unsupervised
- Application
- Cost
- Accuracy
- Latency
- Real-time

HAR Techniques



Action Based Activities

- Gesture Recognition
- Posture recognition
- Behavior Recognition
- Fall Detection
- Activities of Daily Living
- Ambient Assisted Living

Motion Based Activities

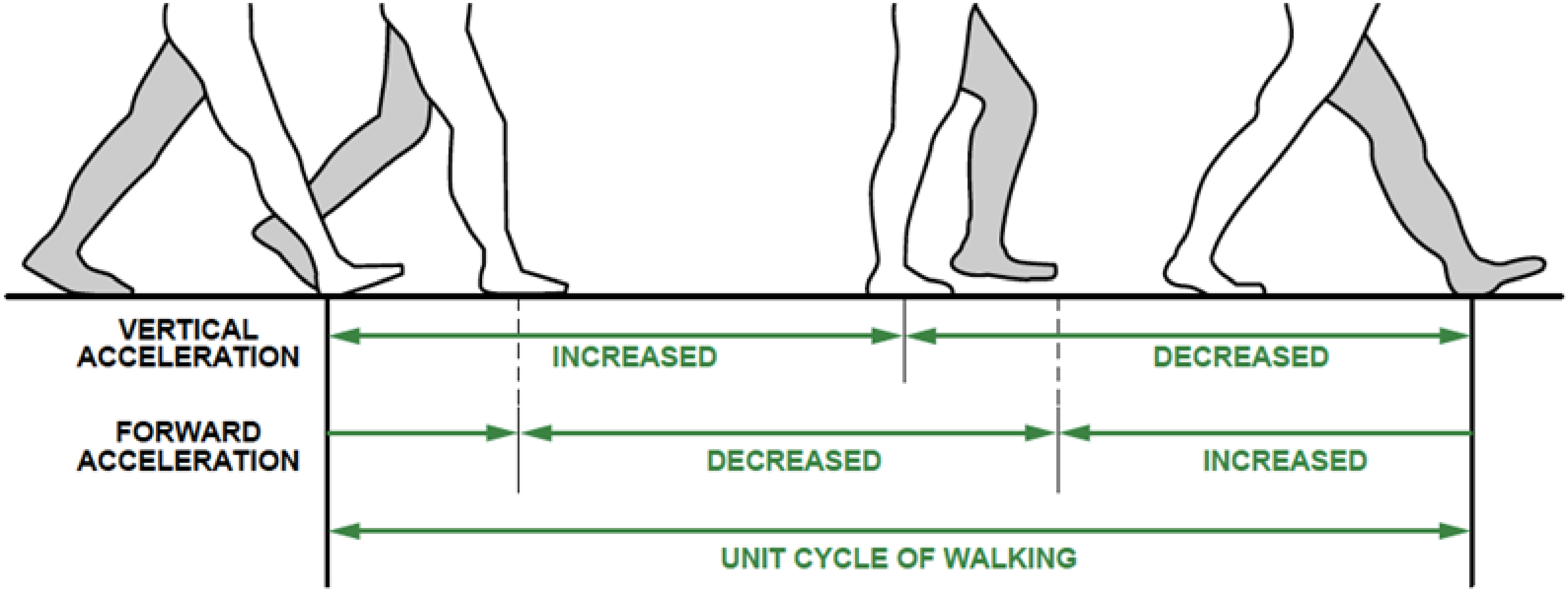
- Tracking
- Motion Detection
- People Counting

Applications of HAR

- Elder Health Care
- Intelligent Environment
- Security and Surveillance
- Human Computer Interaction
- Indoor Navigation
- Shopping Experience

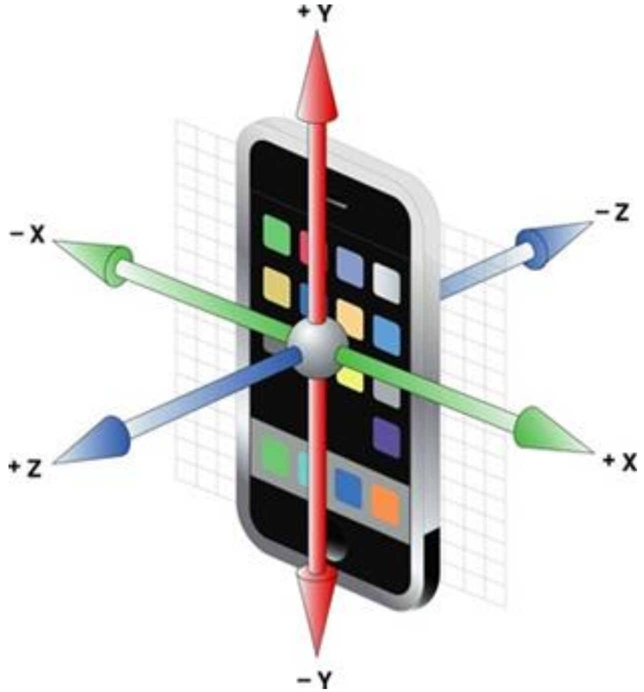
How Do we Monitor Activities?

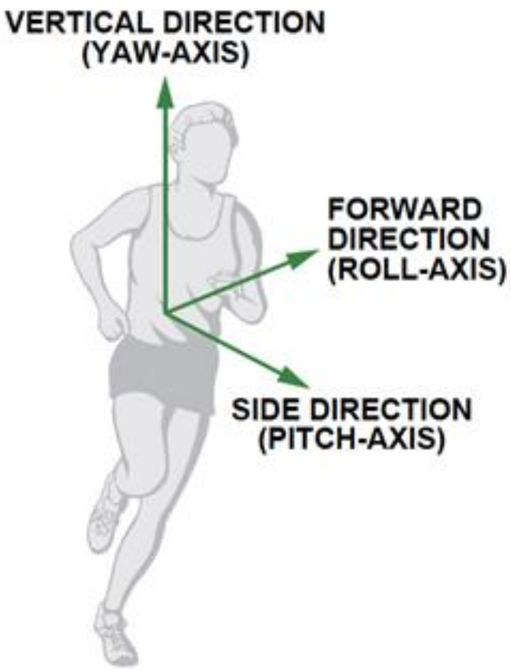
- Activities involve physical movement of body limbs



Inertial Sensors: Accelerometer

- All commodity smartphones have accelerometers
- Measure linear acceleration (m/s^2) in three different directions

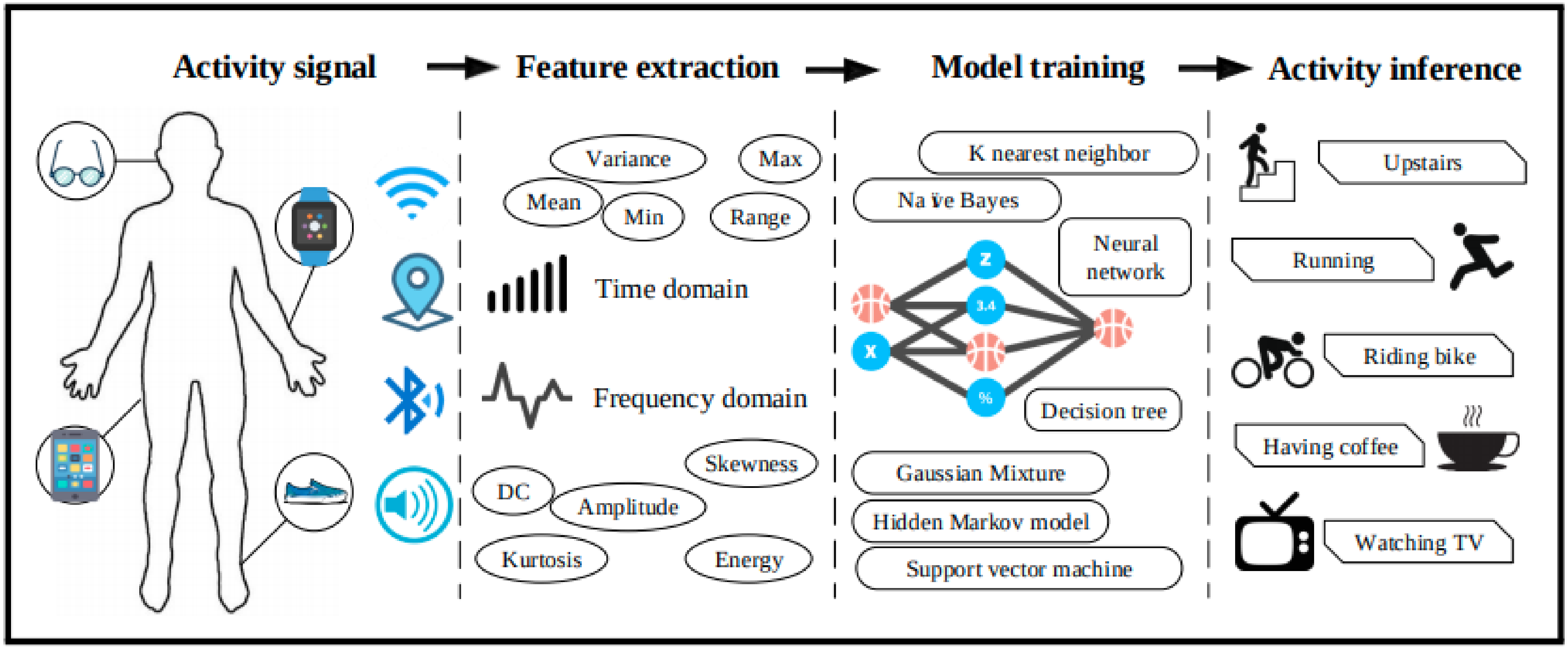




Inertial Sensors:
Gyroscope and Compass

- Gyroscope: measure orientation and angular velocity
- Compass: measure the direction on the earth’s surface toward the north

General Flow of Activity Recognition

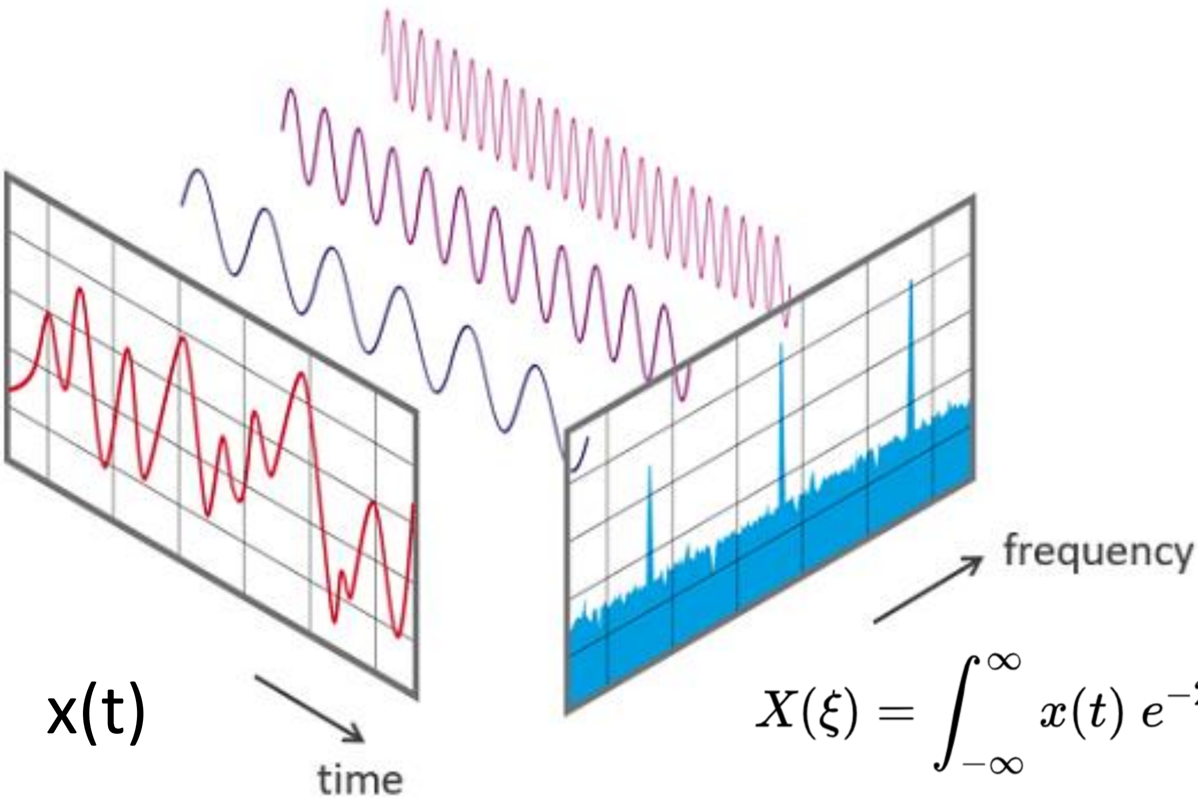


Overview

- ☑ Overview of Activity Recognition
- ☐ **Signal Processing**
  - **Fourier Transformation (FT)**
  - **Filters (Noise Removal)**
  - **Feature Extraction**
- ☐ State-of-the-art Techniques
  - IMU sensor based
    - Gesture Recognition
    - Activity Recognition
  - WiFi signal based Activity Recognition
  - Vision based Activity Recognition
  - Audio based Activity Recognition

Fourier Transformation

- Fourier transformation decomposes a signal into its constituent frequencies.



$$X(\xi) = \int_{-\infty}^{\infty} x(t)\, e^{-2\pi i \xi t}\, dt$$

Why Do we Need Filters?

- Sensor signals often include various noises
- Need to apply various pre-processing techniques to remove noises and highlight the signals we want to capture
- Common pre-processing techniques
  - Moving average filter
  - Exponential filter
  - Median filter
  - Frequency domain filter

Filters (1) - Moving Average Filter

- Use average values of multiple adjacent samples
- Example: Averaging the values for 3 samples
  - Input: x = x1, x2, x3, …, xn where the index is the sample number
  - The output of the moving average filter, s = s1, s2, s3, is:
    - s1 = (x1 + x2 + x3)/3
    - s2 = (x2 + x3 + x4)/3
    - s3 = (x3 + x4 + x5)/3
    - …
    - s(n-2) = (x(n-2) + x(n-1) + xn)/3
- The window size can be different
- The larger the window is, the cleaner the signal becomes
- Too large window may smooth out the important characteristics of the signal (e.g., steps for step detection)

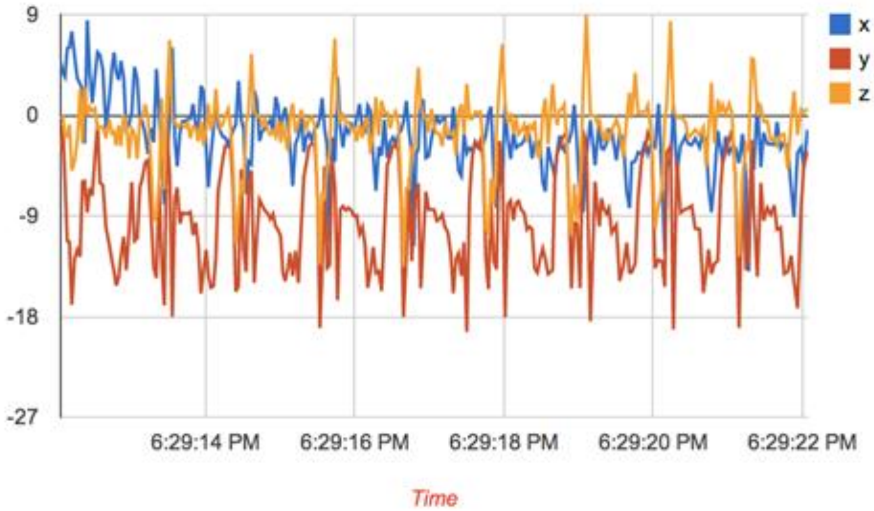
Filter (2) - Exponential Filter

- What if we want to give more weights to recent values?
- The idea in exponential filter is to assign exponentially decreasing weights as the observation get older

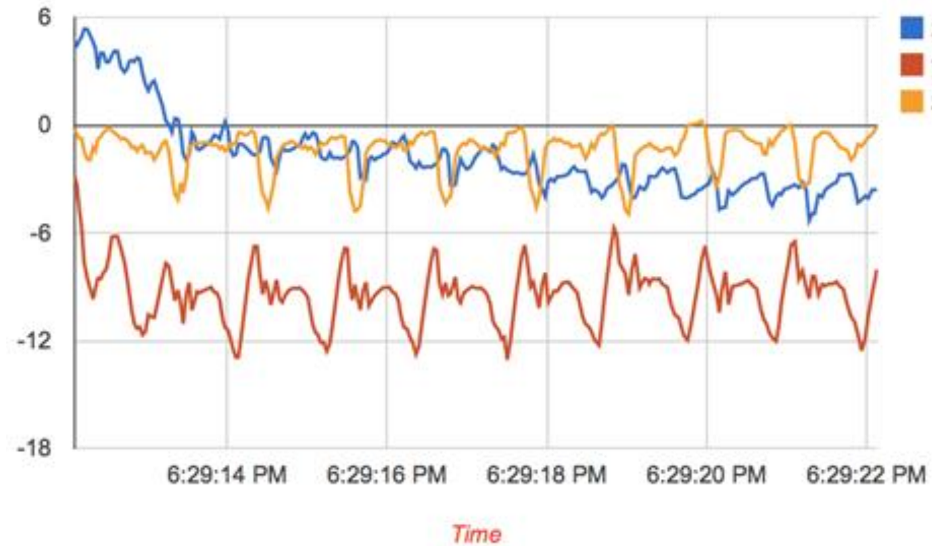
$$s_1 = x_0$$
$$s_t = \alpha x_{t-1} + (1-\alpha)s_{t-1} = s_{t-1} + \alpha(x_{t-1} - s_{t-1}), t > 1$$

- where α is the smoothing factor, and 0 < α < 1
- The filtered output s(t) is a simple weighted average of the current observation x(t) and the previous filtered output s(t-1)

Filter (2) - Effect of Exponential Filter



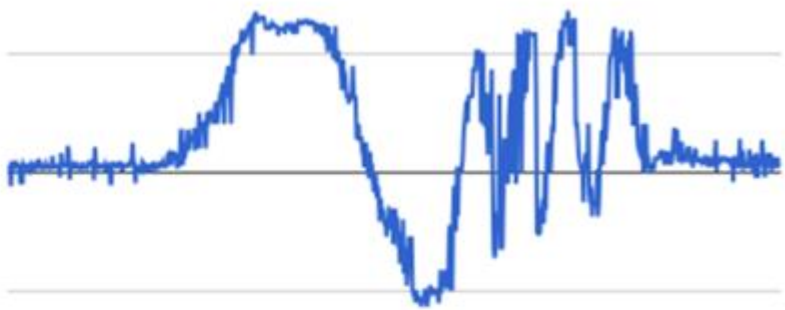
Raw Acceleration Signal



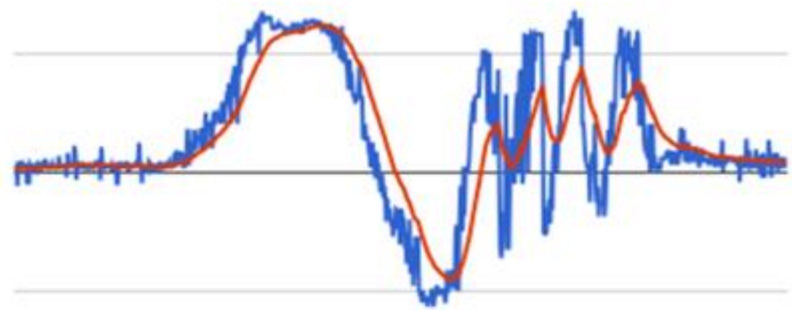
Signal with Exponential Filter (a=1/8)

Filter (2) - Problem of Exponential Filter

- Average out some of the peaks in the data
- Amplitude gets smaller
- Time lag in the peaks, i.e., peaks are slightly shifted to the right



Raw Acceleration Signal

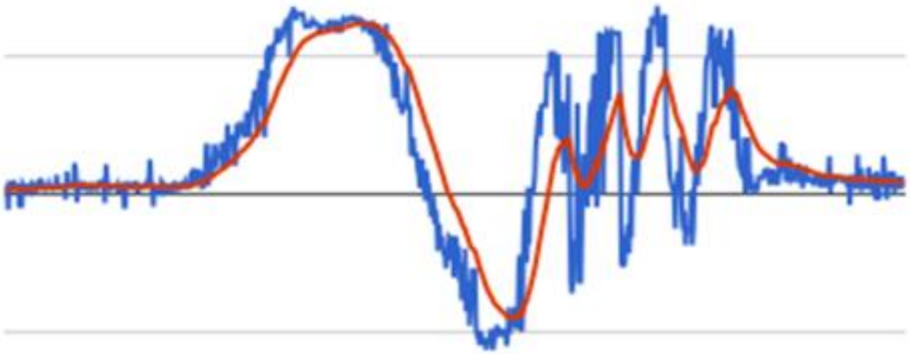


Signal with Exponential Filter

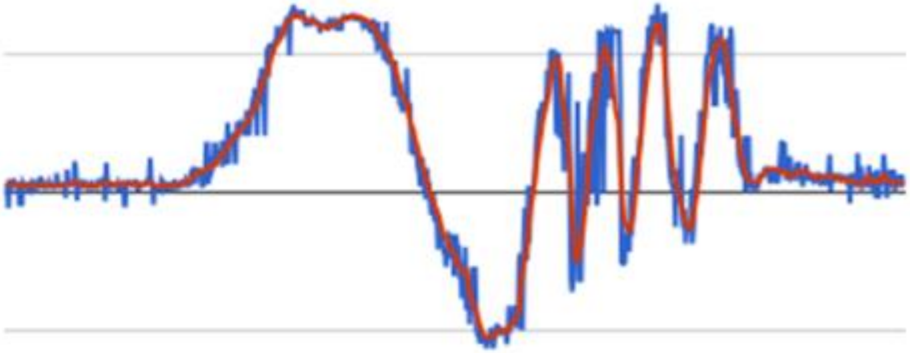
Filter (3) - Median Filter

- Given input accelerometer signal: x = x1, x2, x3, …, xn
- The output of the median filter, s = s1, s2, s3, …, sn, is:
  - s1 = median(x1, x2, x3)
  - s2 = median(x2, x3, x4)
  - s3 = median(x3, x4, x5)
  - …
  - s(n-2) = median(x(n-2), x(n-1), xn)

Filter (3) - Effect of Median Filter



Exponential Filter

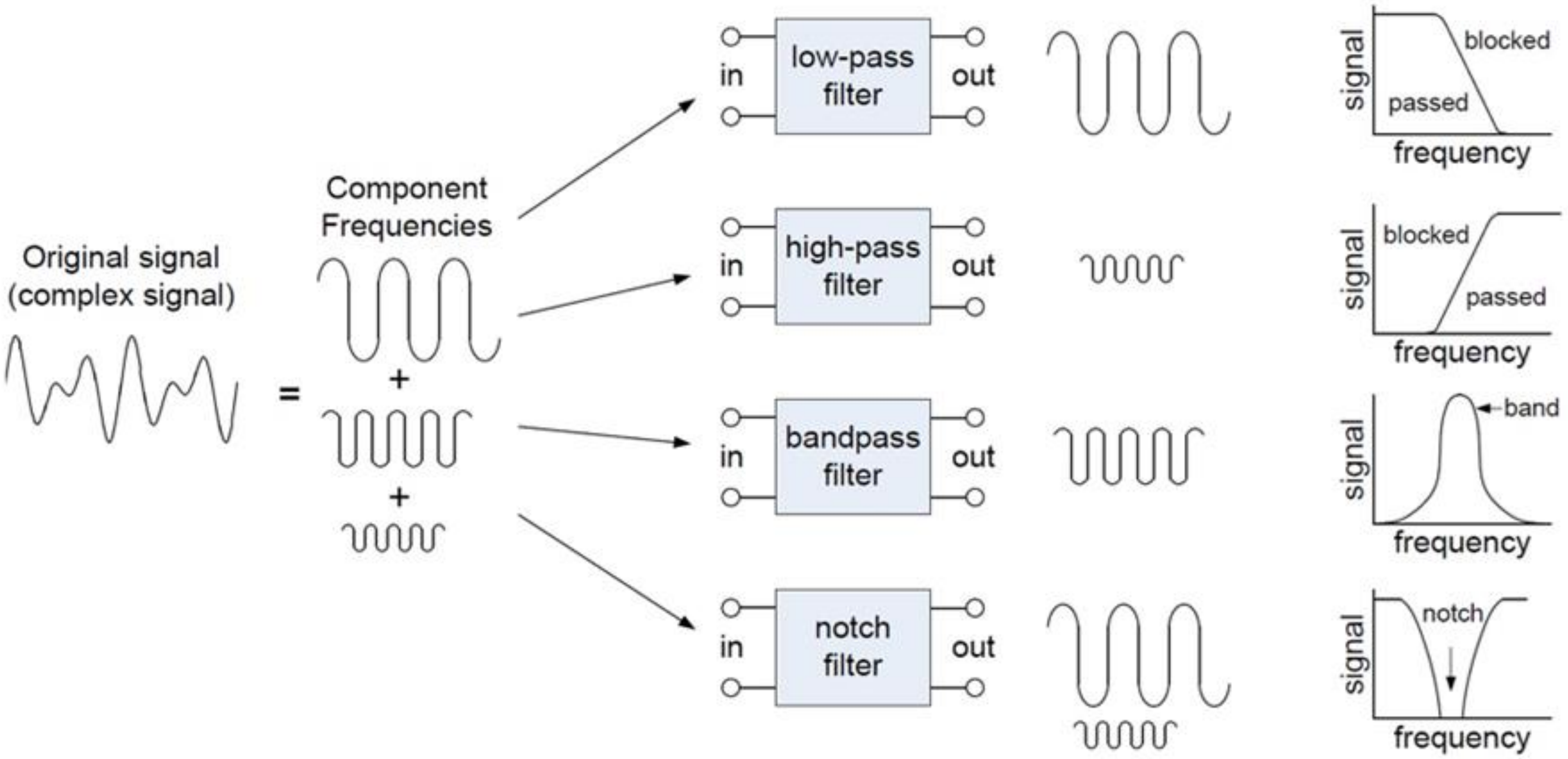


Median Filter

Filter (4) - Frequency Domain Filter

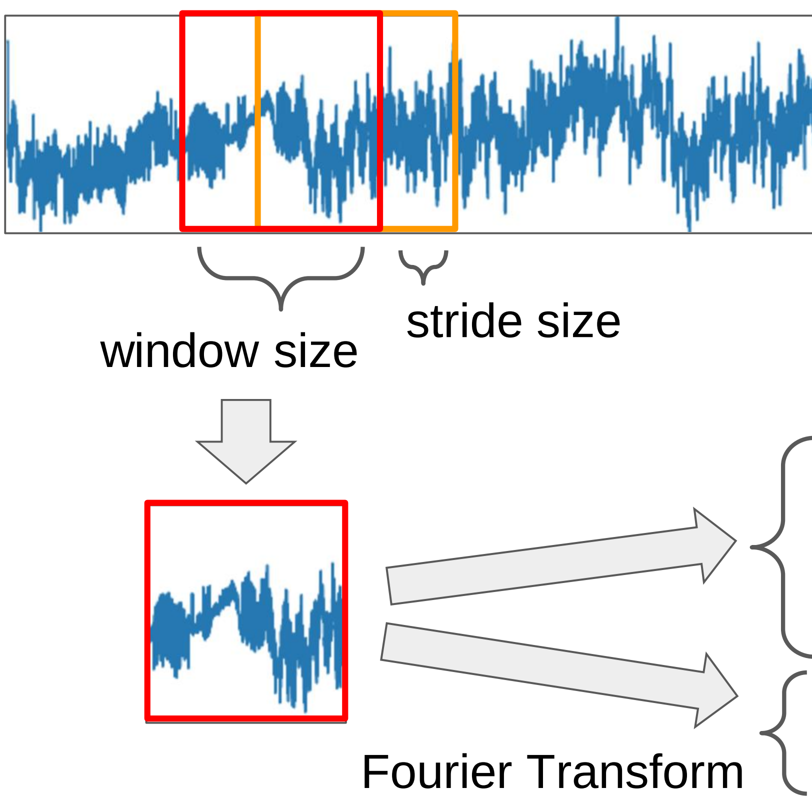
- So far, we have studied the time domain filter
  - First method to remove noises
  - Simple and easy to understand
  - Work well in many practical examples
- In some cases, identifying a good time domain-filter is not easy
- Frequency domain filtering
  - Convert a signal to a weighted sum of sine waves, and remove all the waves whose periods are outside the range that you expect!
  - Jean Baptiste Fourier (1768-1830) proved the mathematical fact that any periodic waveform can be expressed as the sum of an infinite set of sine waves.

Filter (4) - Frequency Domain Filter



Feature Extraction (1) - Sliding Window

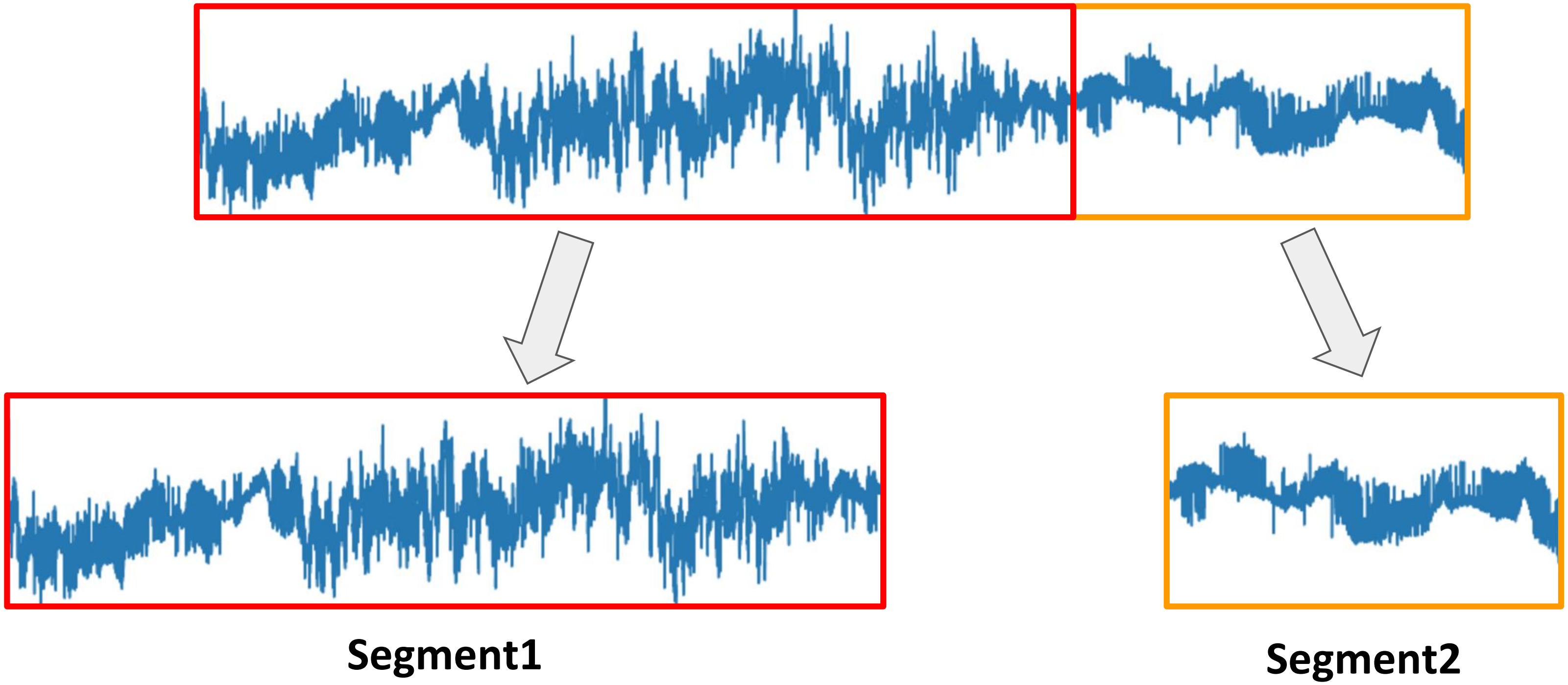
- Extract features with sliding window.
- Reduce the impact of the signal noises.



time	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
max	0.52	...	...	...	...	...	...	...
min	0.21	...	...	...	...	...	...	...
mean	0.41	...	...	...	...	...	...	...
std	0.34	...	...	...	...	...	...	...
entropy	3.3	...	...	...	...	...	...	...
energy	1.4	...	...	...	...	...	...	...

Feature Extraction (2) - Segmentation

- Split segments
- Extract features with segments
- Classify each segment



Feature Extraction (3)

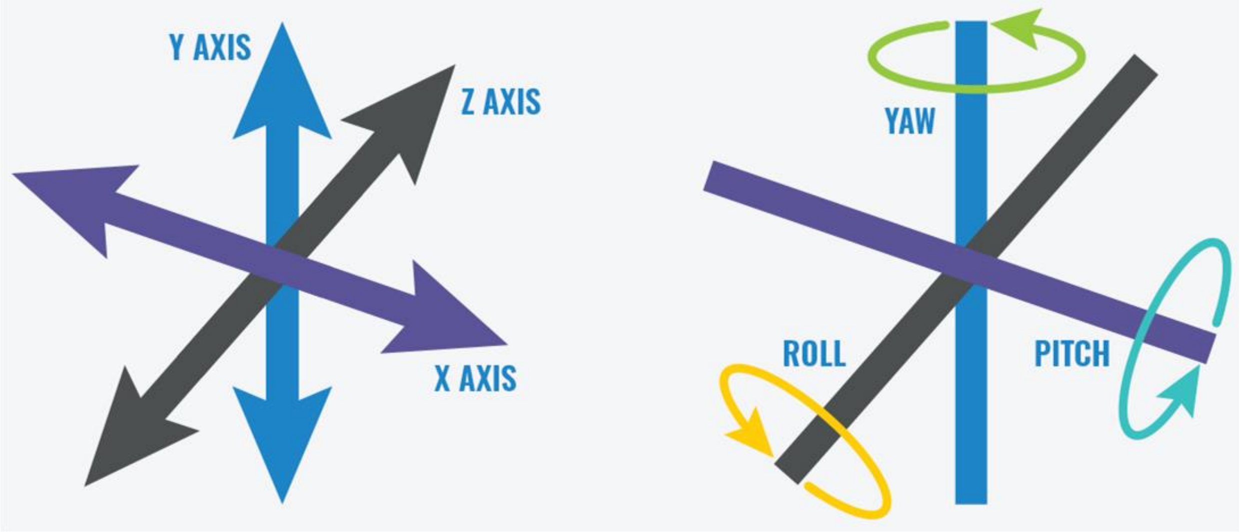
- Time Domain Features
  - Mean
  - Standard deviation
  - Maximum
  - Minimum
  - Cross-correlation
  - RMS
- Frequency Domain Features
  - Energy
  - Entropy
  - Coefficient sum
  - Dominant frequency

Overview

- ☑ Overview of Activity Recognition
- ☑ Signal Processing
  - Fourier Transformation (FT)
  - Filters (Noise Removal)
  - Feature Extraction
- ☐ **State-of-the-art Techniques**
  - **IMU sensor based**
    - **Gesture Recognition**
    - **Activity Recognition**
  - **WiFi signal based Activity Recognition**
  - **Vision based Activity Recognition**
  - **Audio based Activity Recognition**

IMU sensor

- Accelerometer
- Gyroscope
- + Magnetometer

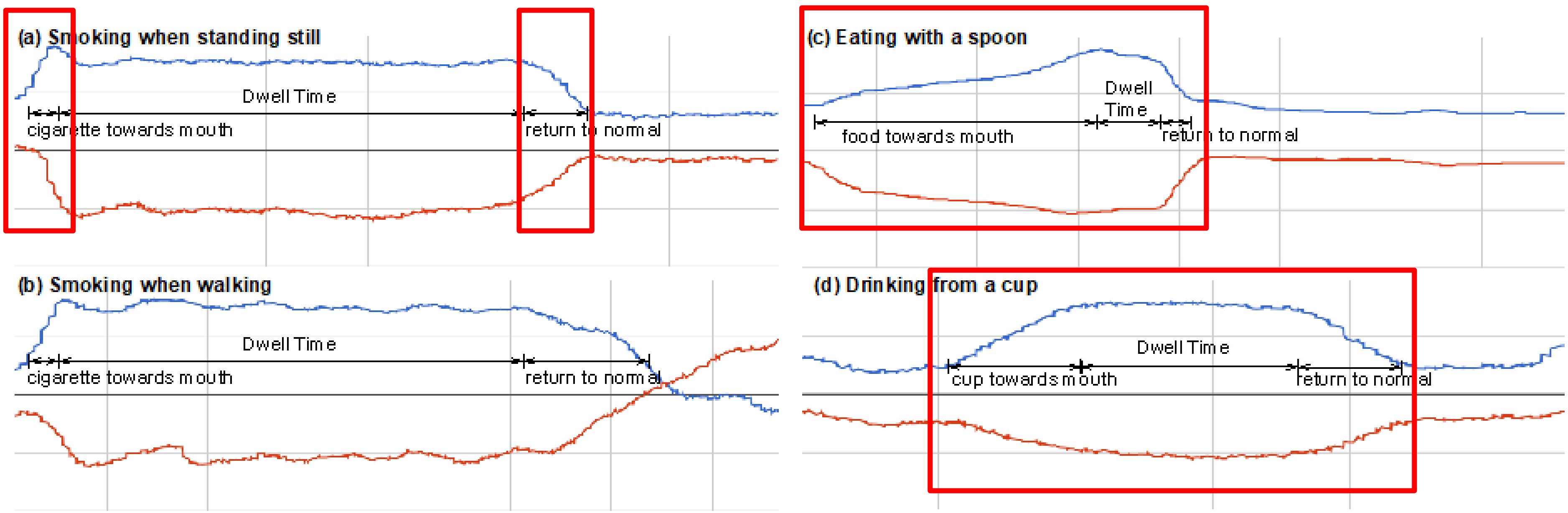


IMU-based Gesture Recognition

- IMU Sensors on a Wristband
- 3 Classes: Smoking, Eating, Other
- Using Segmentation

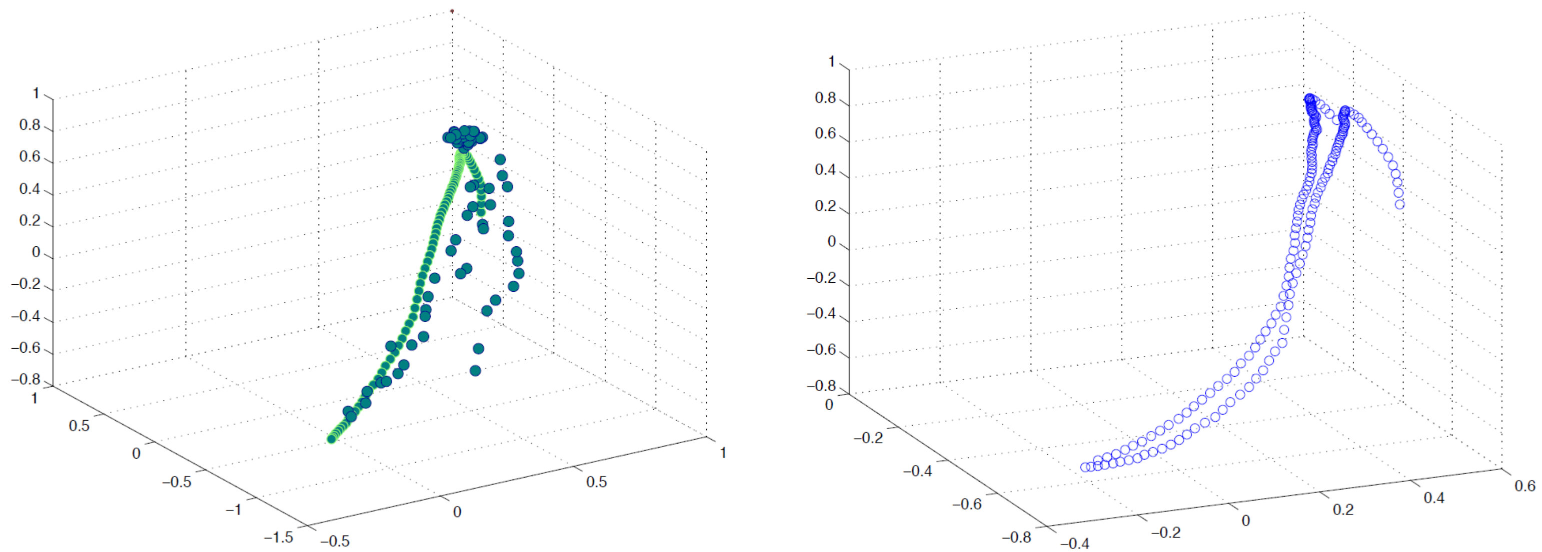
RisQ: Recognizing Smoking Gestures with Inertial Sensors on a Wristband, Mobisys 2014

IMU-based Gesture Recognition



- Quick change in orientation when taking a cigarette
- Long dwell time

IMU-based Gesture Recognition



- 10sec sliding window trajectory
- Check rapid increase
- Detect peak

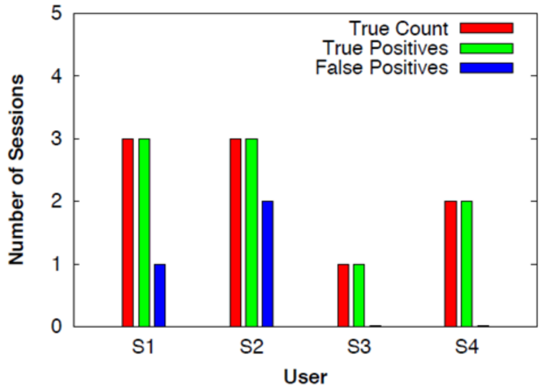
IMU-based Gesture Recognition

- Feature Extracting with a Segment
  - Duration features
  - Velocity features
  - Displacement features
  - Angle features
- Classification
  - Random Forest
  - Conditional Random Field

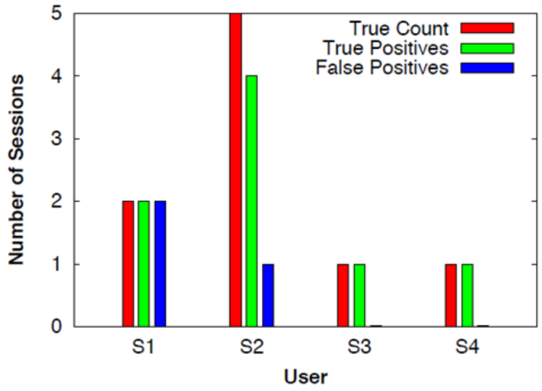
	Accuracy	Recall	Precision	False-positive rate
RF	93.00%	0.85	0.72	0.023
CRF	95.74%	0.81	0.91	0.005

IMU-based Gesture Recognition

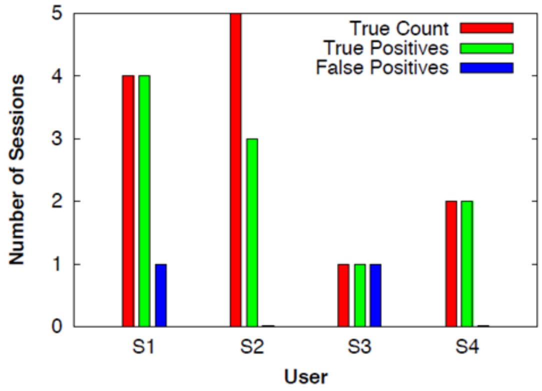
- User Study in the Wild
- Monitoring App
- Wristband



(a) Day 1



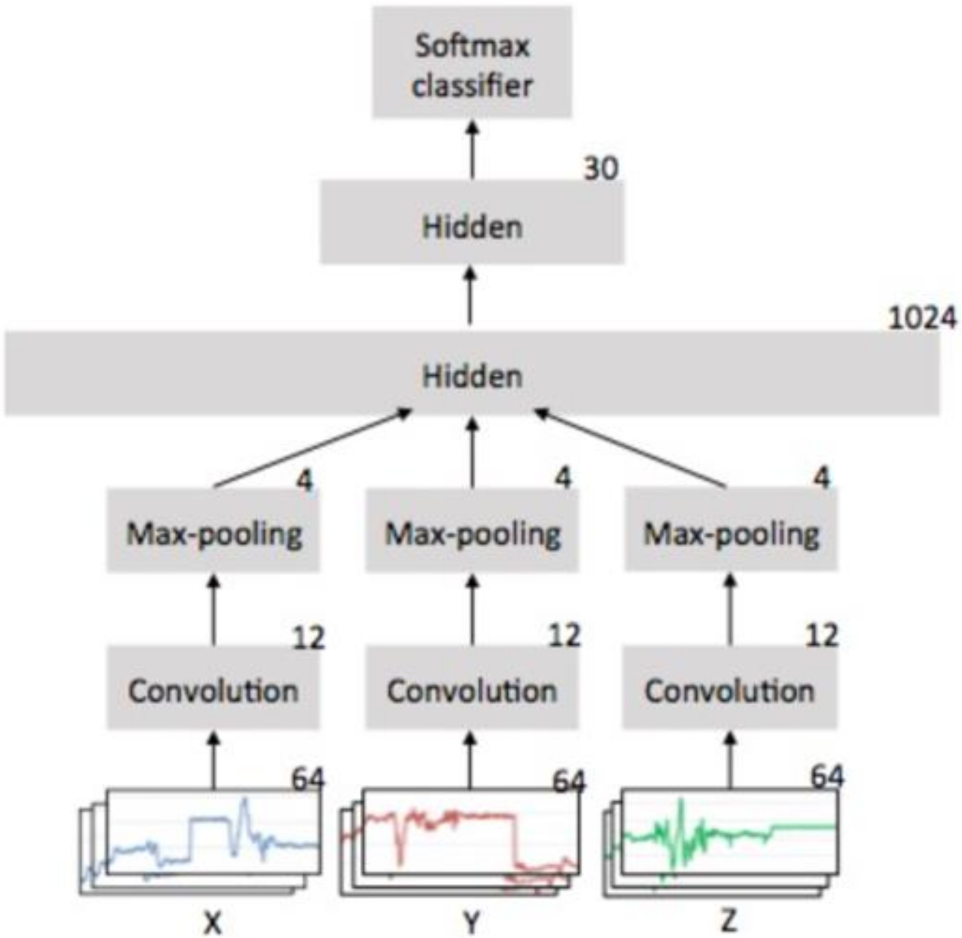
(b) Day 2



(c) Day 3

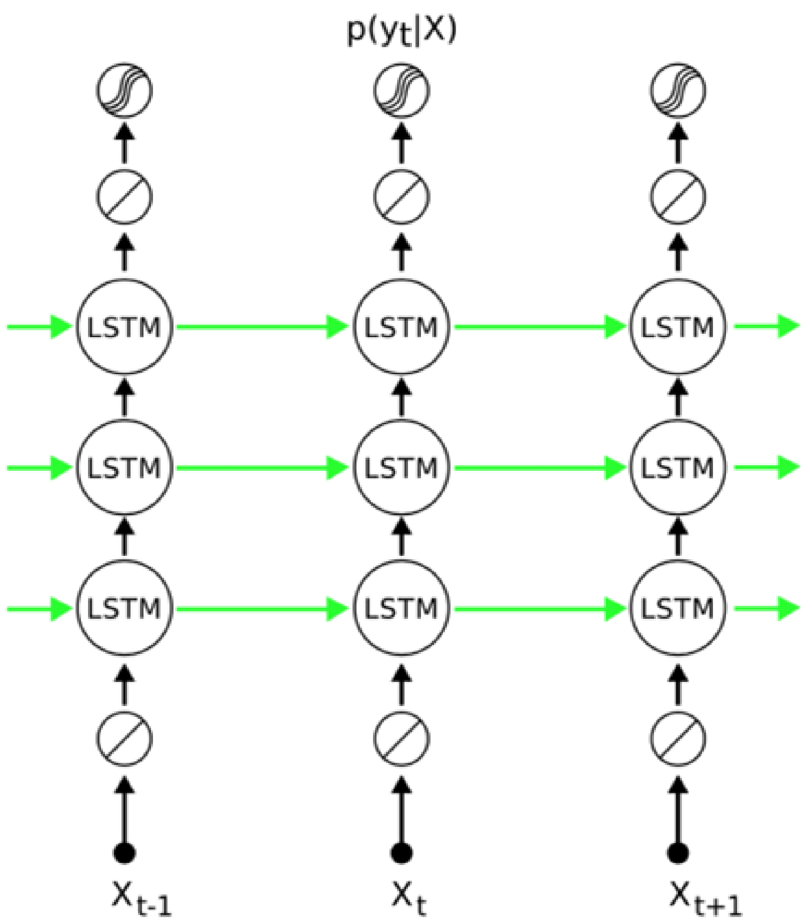
IMU-based HAR (CNN)

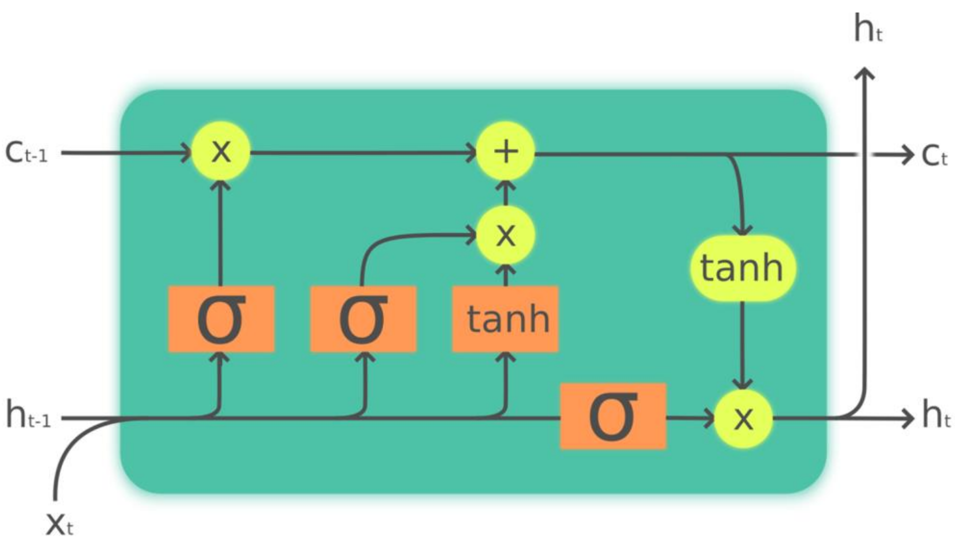
- CNN for frame-based HAR (2014)



IMU-based HAR (LSTM)

- Deep LSTM Networks for HAR (2016)

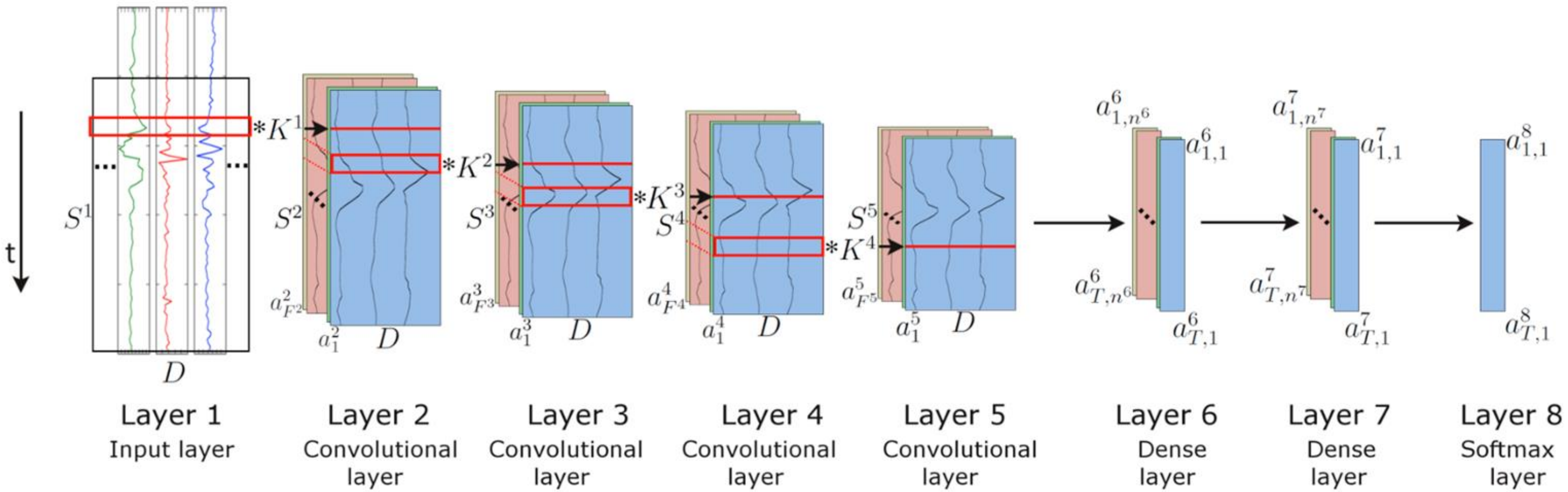




LSTM

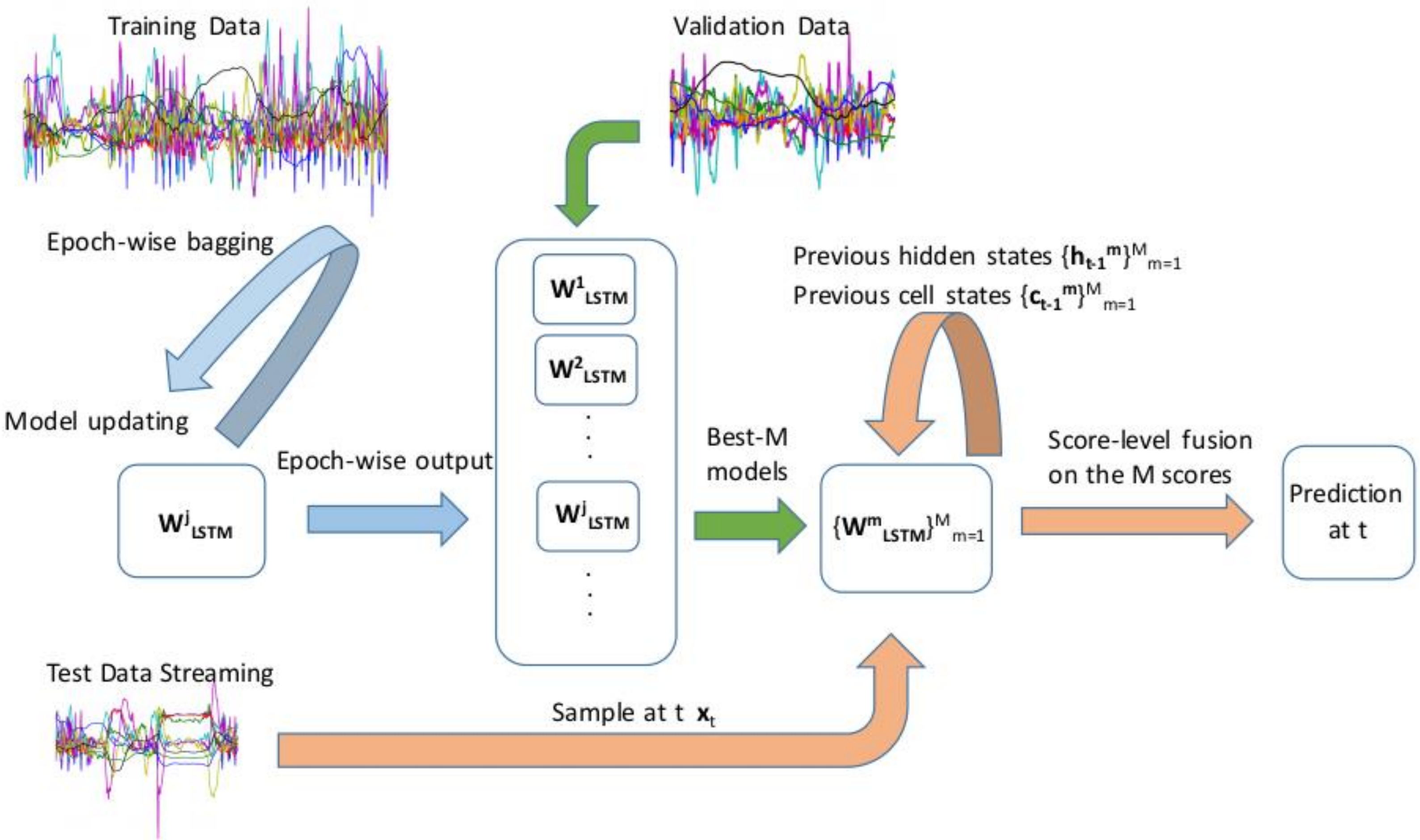
IMU-based HAR (CNN + LSTM)

- Combinations of convolutional layers and LSTM (2016)
- 1D convolutional layers extract features.



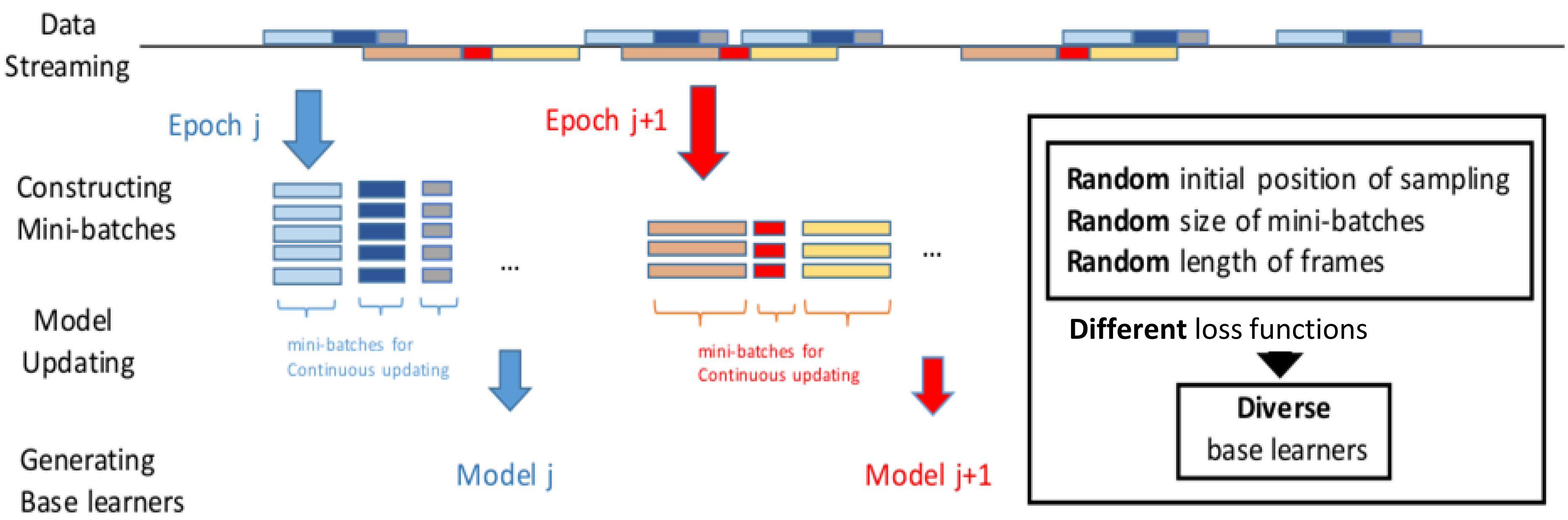
IMU-based HAR (LSTMs) (1/2)

- Ensembles of Deep LSTM Learners. (2017)



IMU-based HAR (LSTMs) (2/2)

- Multiple LSTM learners trained with random mini-batches.

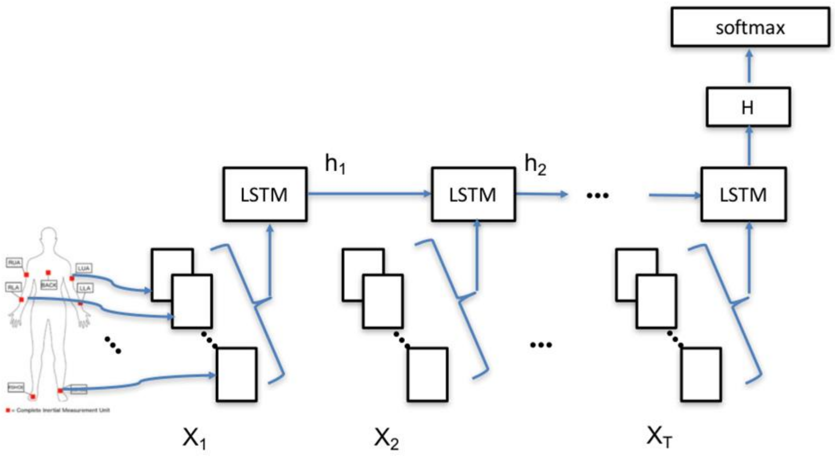


Attention based LSTM (1/7)

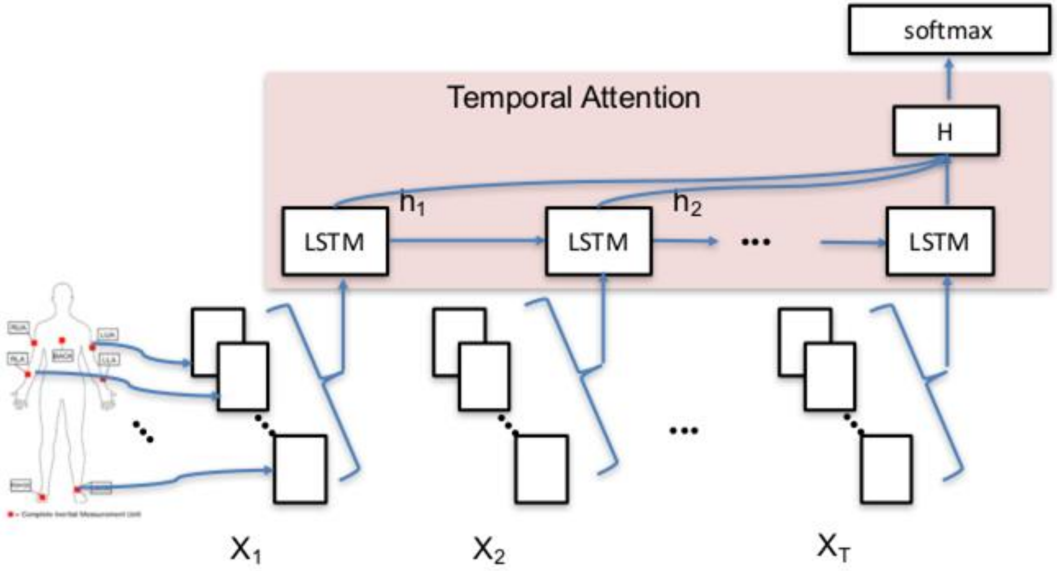
- *Understanding and Improving Recurrent Networks for Human Activity Recognition by Continuous Attention, 2018*
- Limitation of LSTM: Even though LSTM is designed to mitigate the long-distance dependencies problem, is it hard to guarantee that we will learn to handle those properly.
  ⇒ Attention based LSTM

Attention based LSTM (2/7)

Standard LSTM



Temporal attention based LSTM



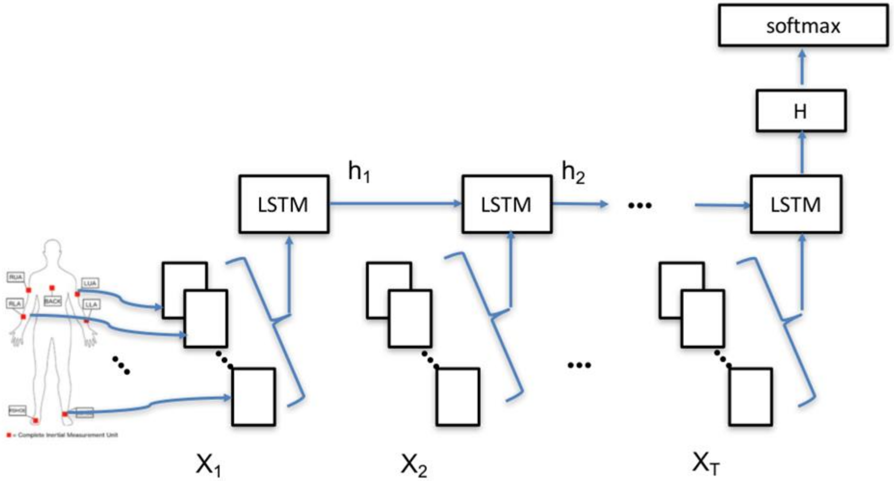
$$\mathbf{H} = \sum_{t=1}^{T} \alpha_t \mathbf{h}_t$$

$$\alpha_t = \frac{\exp\{\text{score}(\mathbf{h}_T, \mathbf{h}_t)\}}{\sum_{s=1}^{T} \exp\{\text{score}(\mathbf{h}_T, \mathbf{h}_s)\}}$$

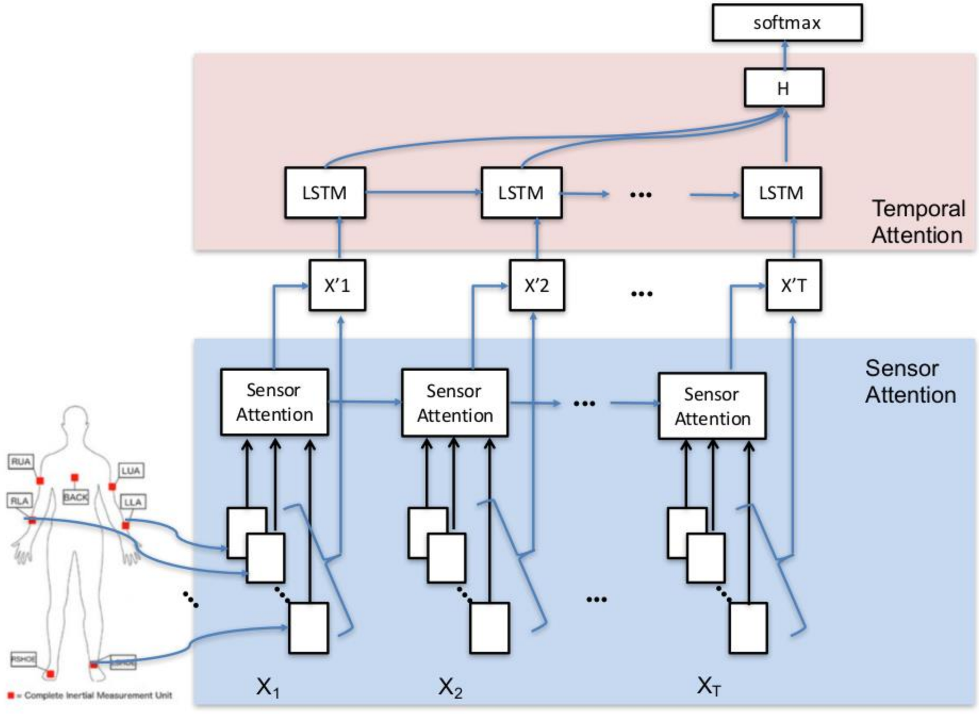
$$\text{score}(\mathbf{h}_t, \mathbf{h}_s) = \mathbf{h}_t^T \mathbf{W}_\alpha \mathbf{h}_s$$

Attention based LSTM (3/7)

Standard LSTM



Attention-based LSTM
(Temporal Attention + Sensor Attention)



Attention based LSTM (4/7)

- Continuous temporal attention regularization to encourage continuity.

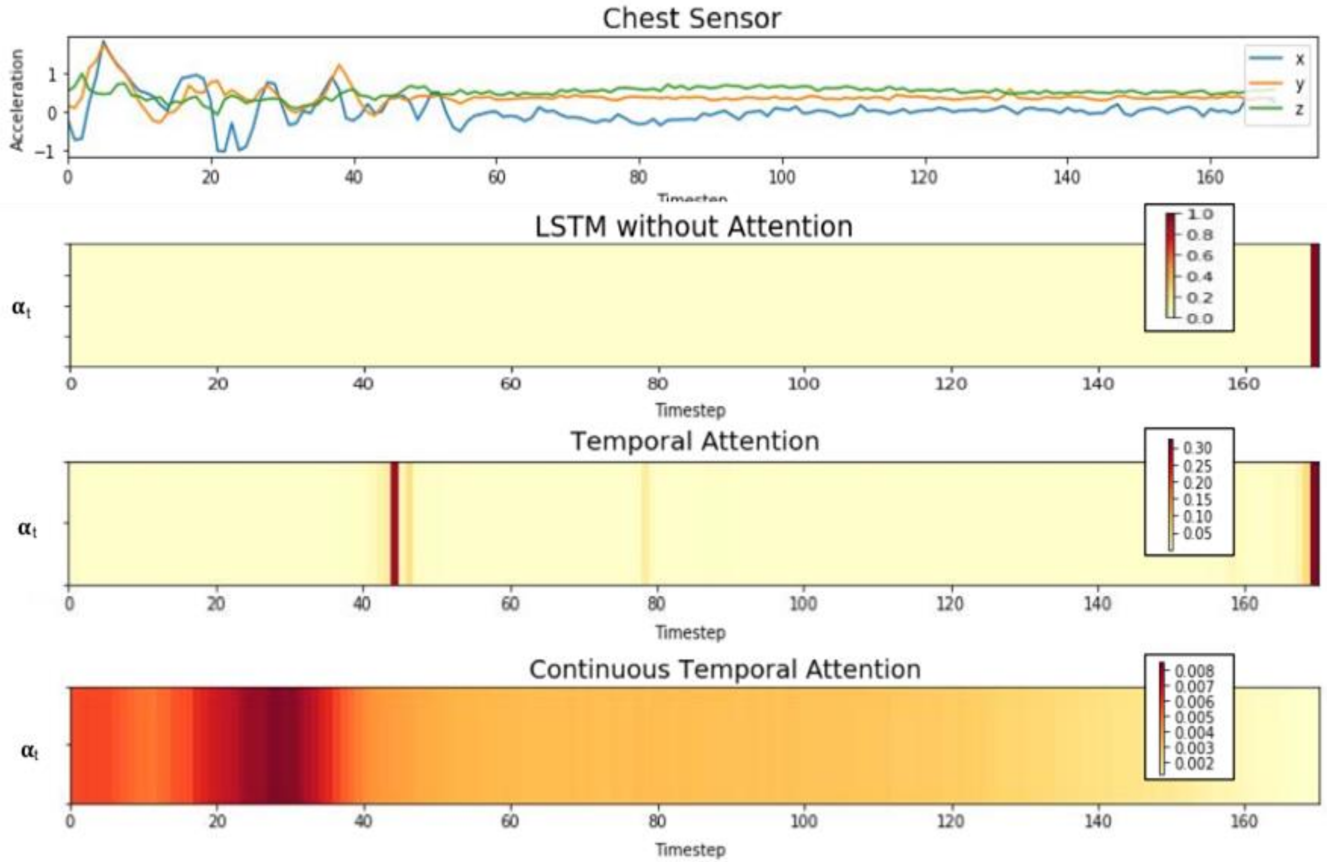
$$\Omega_T(\boldsymbol{\alpha}) = \lambda_1 \sum_t |\alpha_t - \alpha_{t-1}|$$

- Continuous sensor attention regularization to discourage transitions.

$$\Omega_S(\boldsymbol{\beta}) = \lambda_2 \sum_t |\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1}|$$

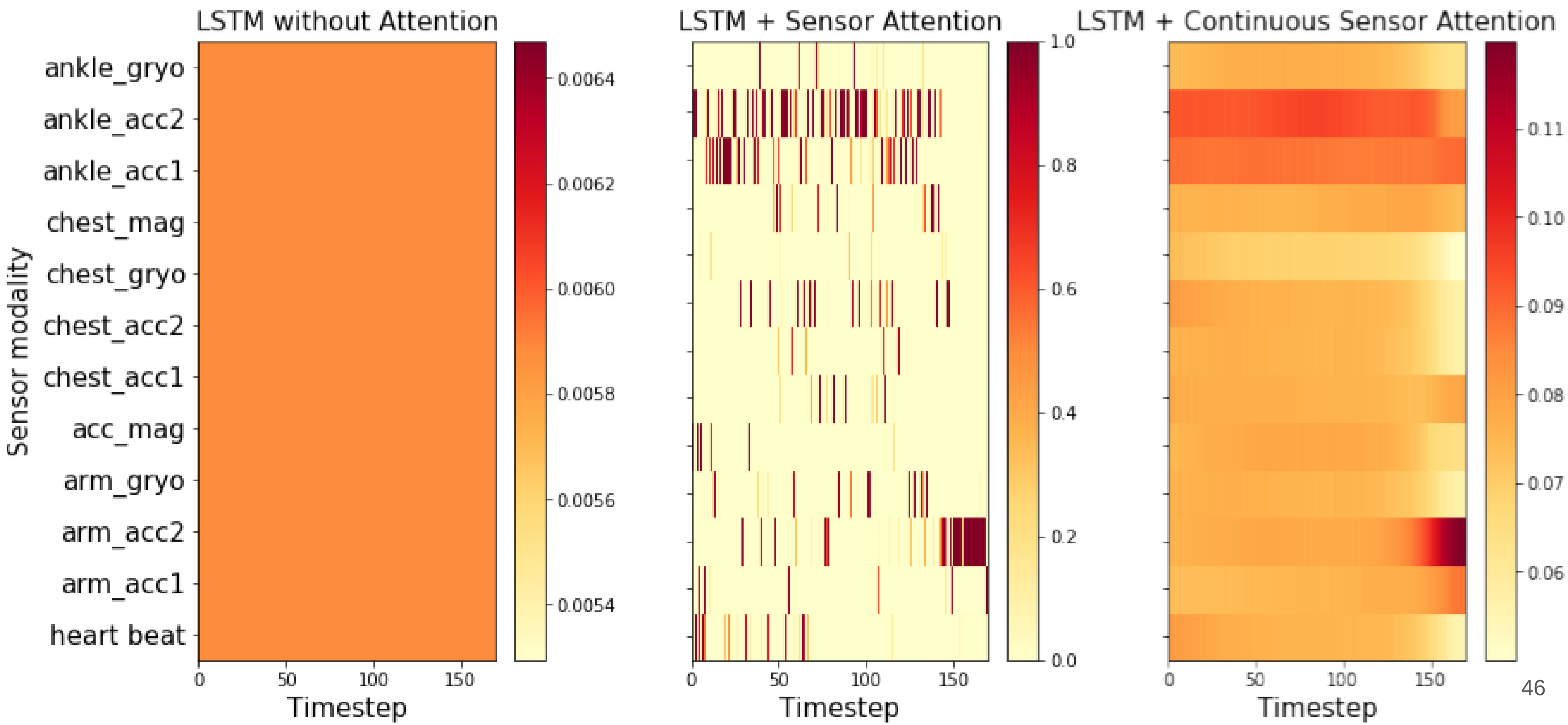
Attention based LSTM (5/7)

- Attention weights of models with and without **temporal attention**.



Attention based LSTM (6/7)

- Attention weights of models with and without **sensor attention**.



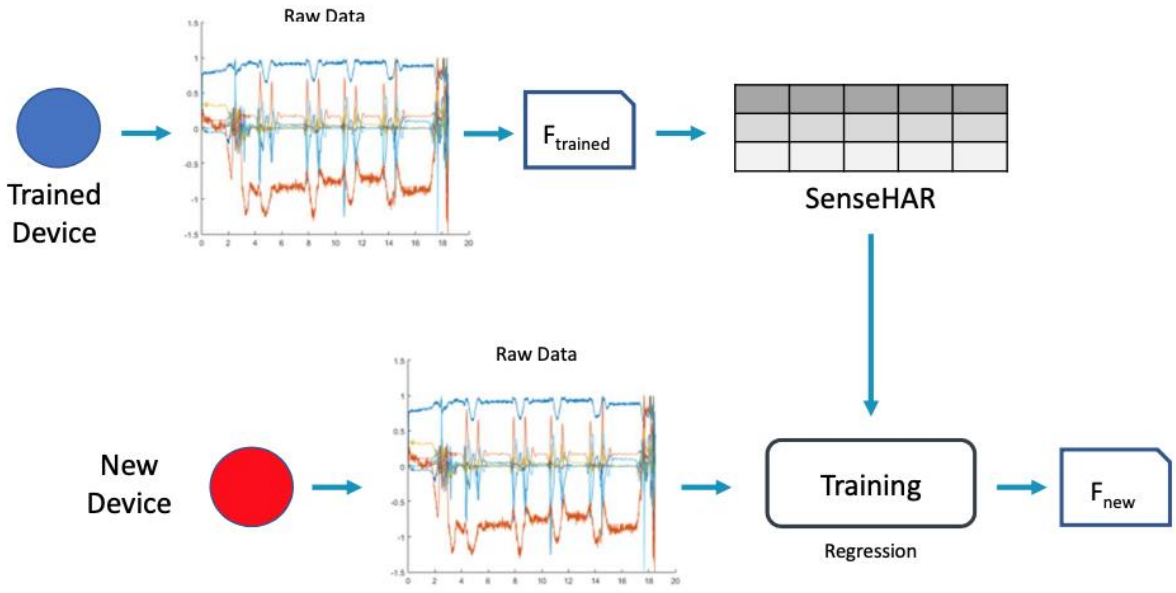
Attention based LSTM (7/7)

Models	**PAMAP2**	**DG**	**Skoda**[4]
LSTM baseline ([12]) (LSTM without Attention)	0.7548	0.6675	0.9040
DeepConvLSTM ([19])	0.7480	0.7344[3]	0.9120
LSTM-S ([11])	0.8820	0.7600	0.9210
LSTM + Temporal Attention	0.8052	0.7913	0.9240
LSTM + Sensor Attention	0.7384	0.6700	0.9002
LSTM + Continuous Temporal Attention	0.8629	0.8216	**0.9381**
LSTM + Continuous Sensor Attention	0.7797	0.7817	0.8802
LSTM + Continuous Temporal + Continuous Sensor Attention[5]	**0.8996**	**0.8373**	0.8903

mean F1 score

SenseHAR: Virtual Activity Sensor

- Data variances from different kinds of wearable/mobile devices.
- Mapping different kinds of raw sensor values to the same feature space.



SenseHAR: Virtual Activity Sensor

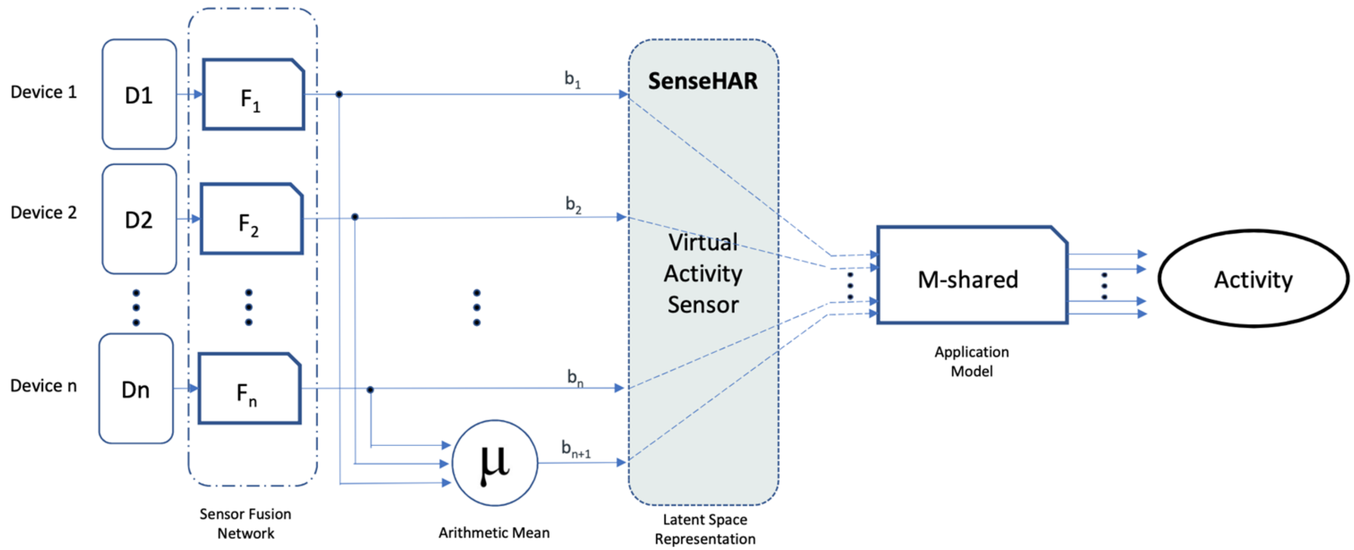


Figure 2: Branched Training Framework to construct SenseHAR: It has (a)Sensor Fusion Network comprising 'n' Sensor fusion models which maps to the shared latent space 'SenseHAR' (b) Shared Application model to predict activities.

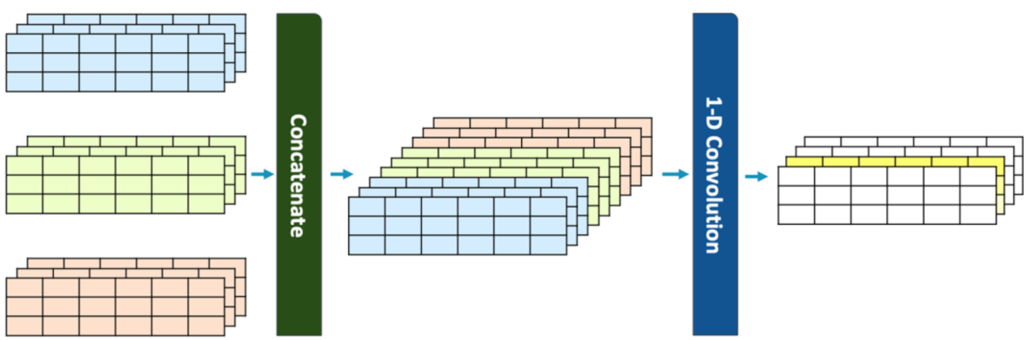


Figure 4: Sensor Fusion Model - Stage 2: Captures the correlation across the corresponding axes in different sensors.

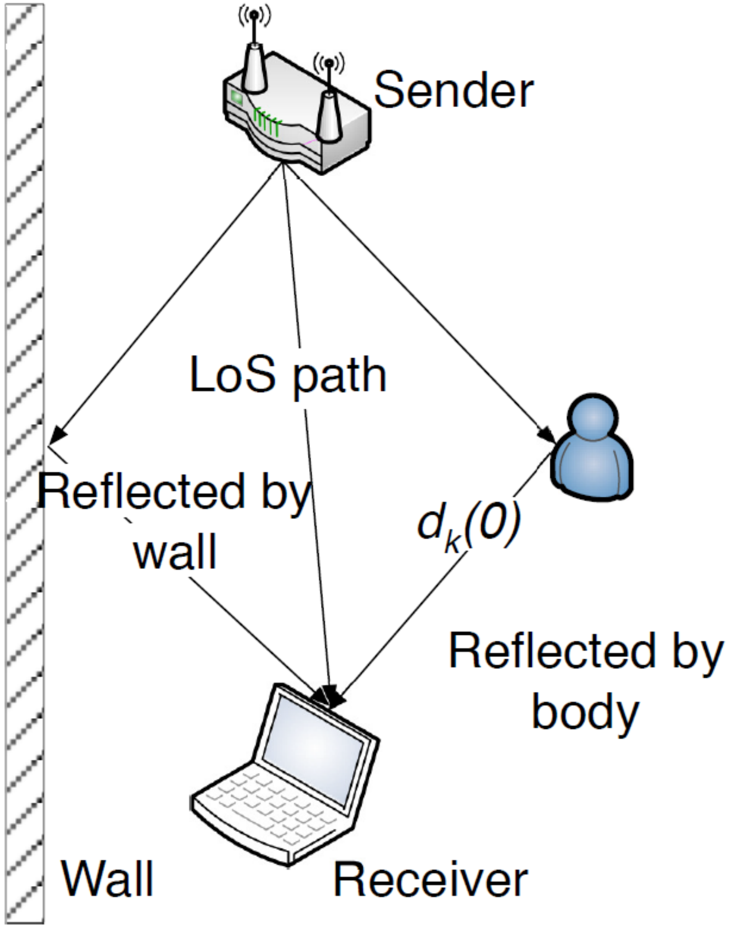
WiFi signal based HAR

- WiFi signals are available almost everywhere and they are able to monitor surrounding activities

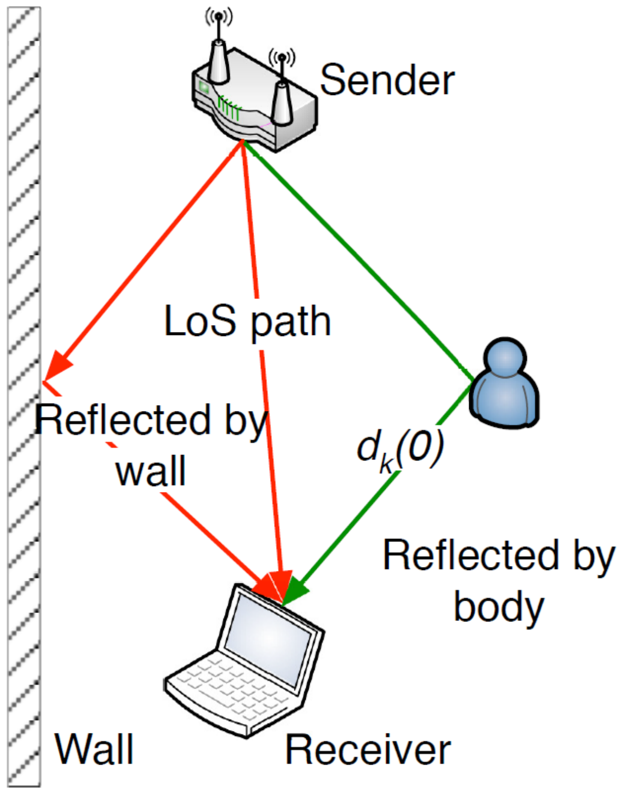
Understanding and Modeling of WiFi Signal Based Human Activity Recognition, Mobicom 2015

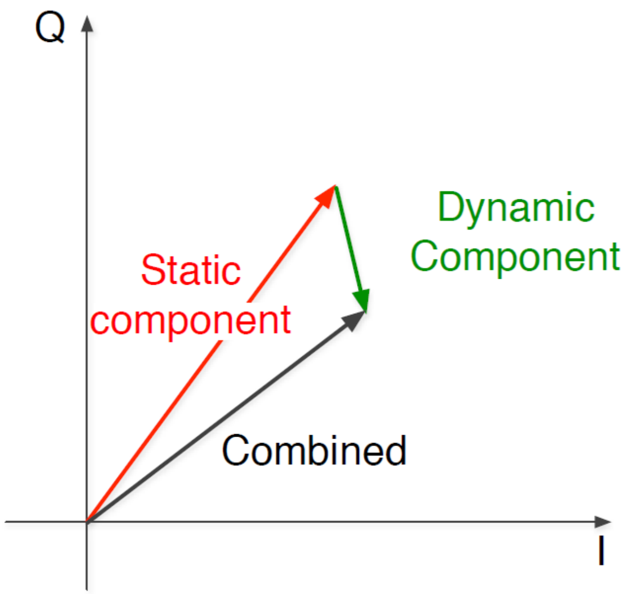
WiFi signal based HAR

- Multipaths contain both static component and dynamic component
- Each path has different phase
- Phases determine the amplitude of the combined signal

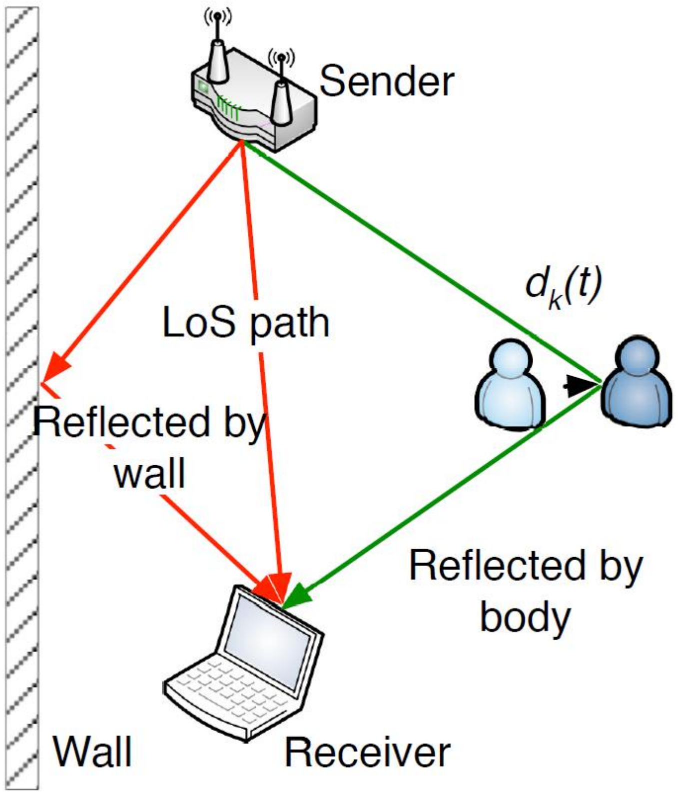


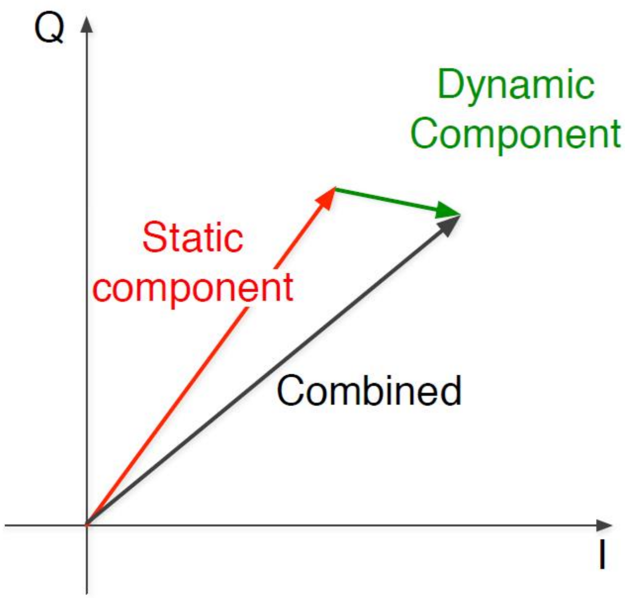
WiFi signal based HAR





WiFi signal based HAR

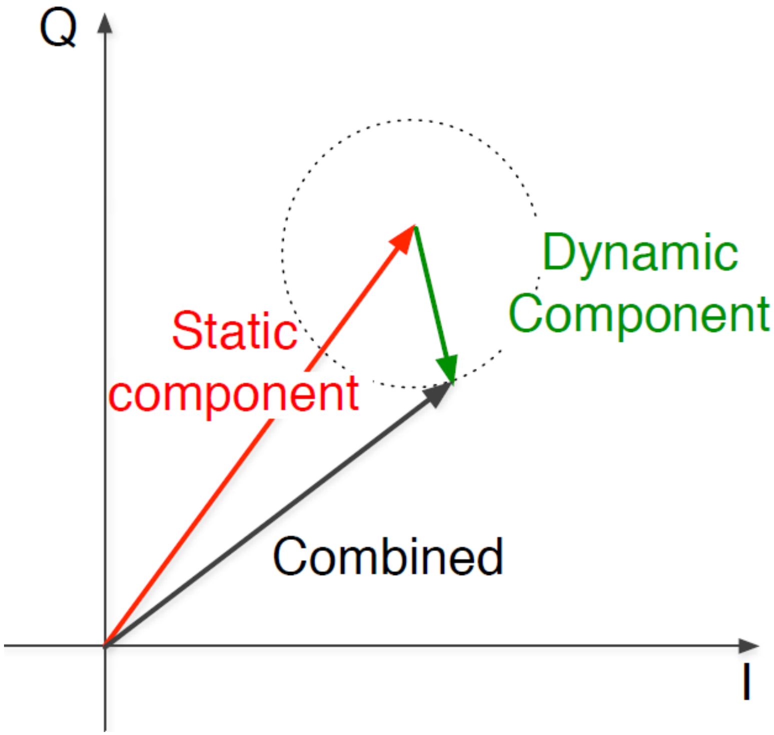




WiFi signal based HAR

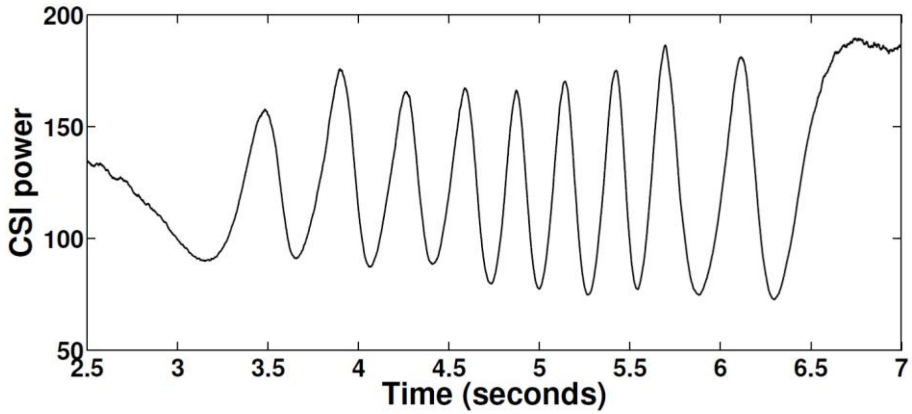
Interpreting CSI amplitude

- Phases of paths are determined by path length
- Path length change of one wavelength gives phase change of 2π
- Frequency of amplitude change can be converted to movement speed

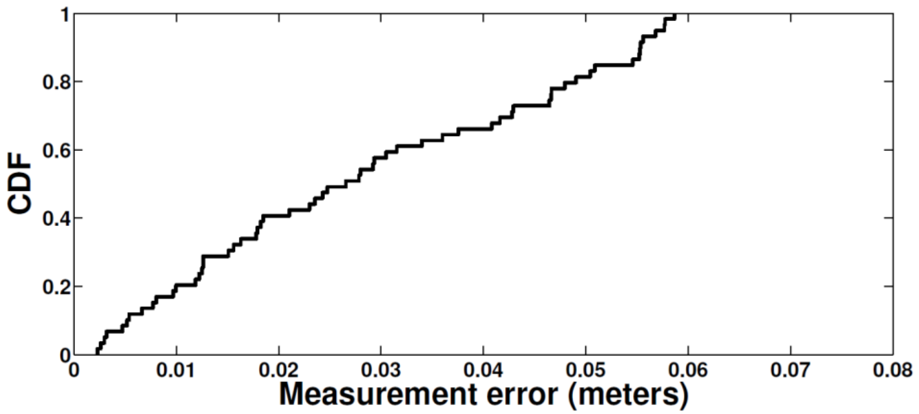


WiFi signal based HAR

- How accurate is it?
  - Wave length → 5~6cm in 5GHz band
  - Average distance measurement error of 2.86cm



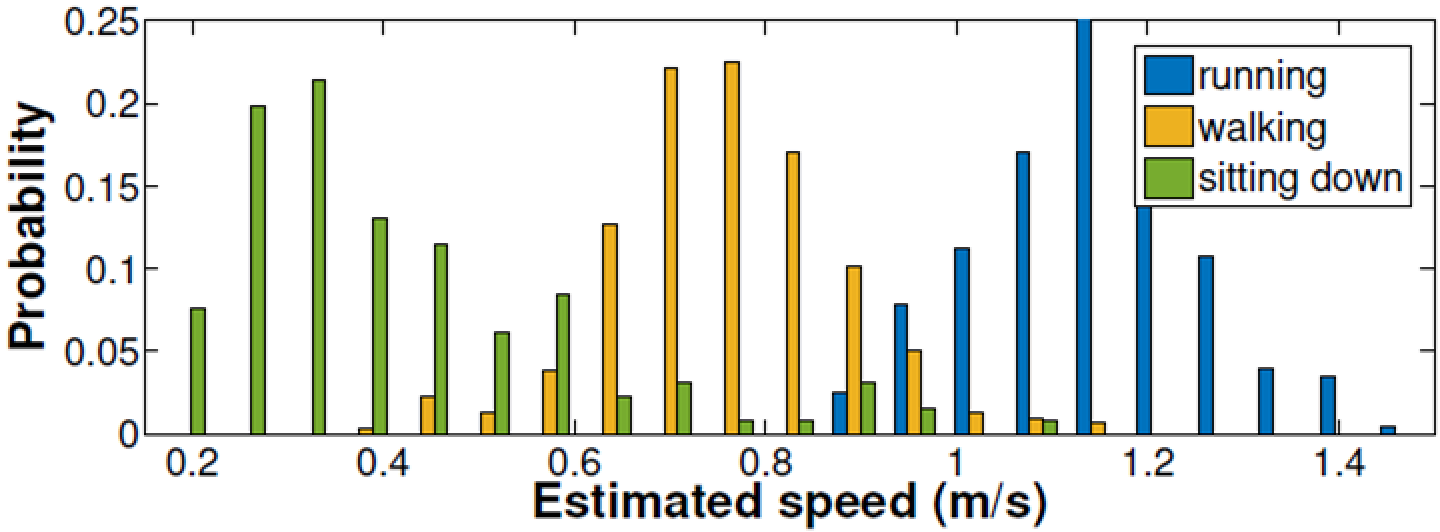
Waveform with regular moving speed



Moving distance measurement error

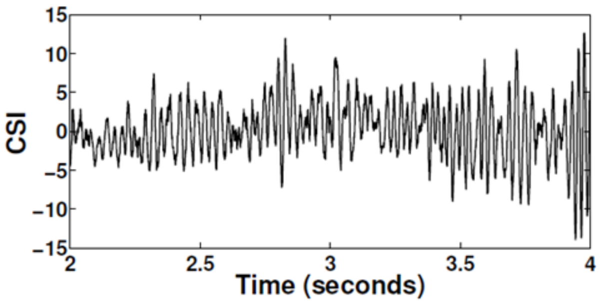
WiFi signal based HAR

- How robust is it?
  - Robust over different multipath conditions and movement directions
  - Linear combination of multipath do not change frequency

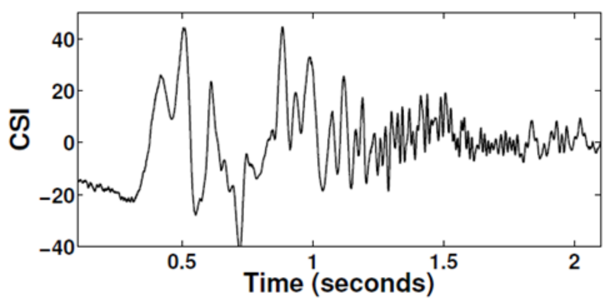


WiFi signal based HAR

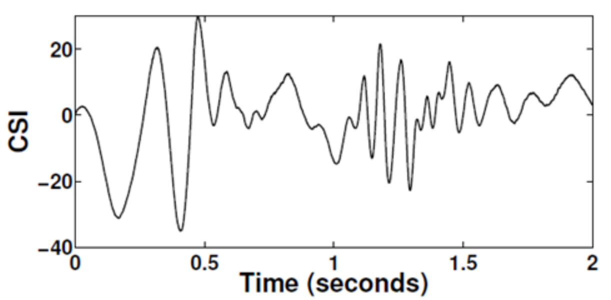
- Activities are characterized by
  - Movement speeds
  - Change in movement speeds
  - Speeds of different body components



Walking



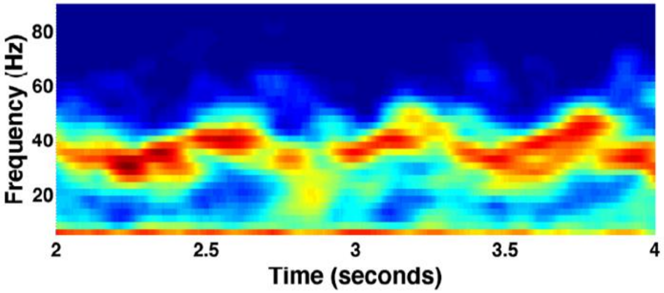
Falling



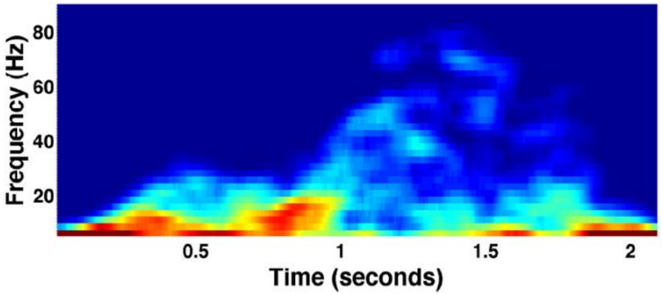
Sitting down

WiFi signal based HAR

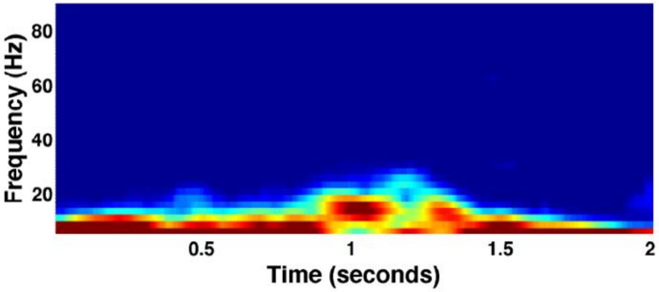
- Use time-frequency analysis to extract features
- Use HMM to characterize the state transitions of movements



Walking



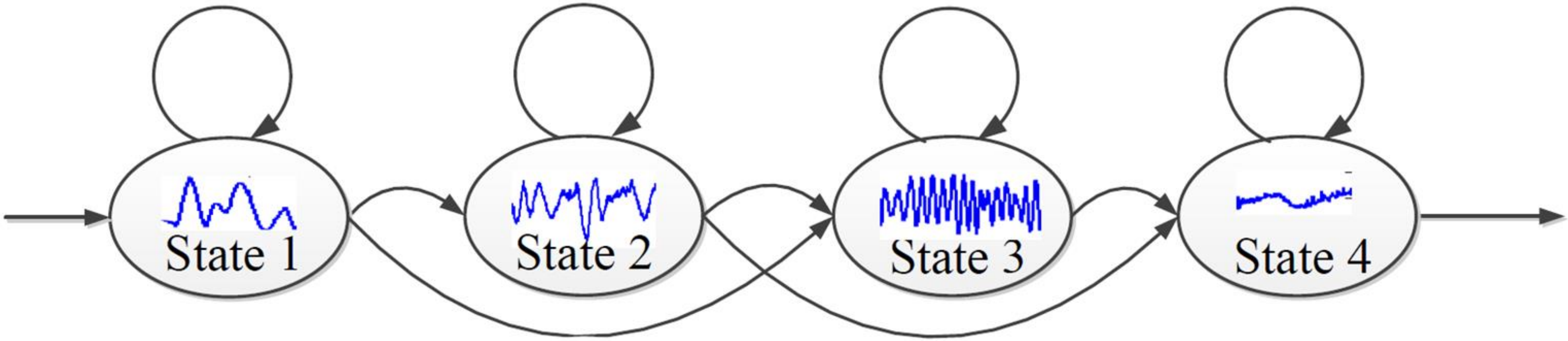
Falling



Sitting down

WiFi signal based HAR

- Build one HMM model for each activity
- Determine states based on observations in waveform patterns
- State durations and relationships are captured by transition probabilities



Vision-based HAR

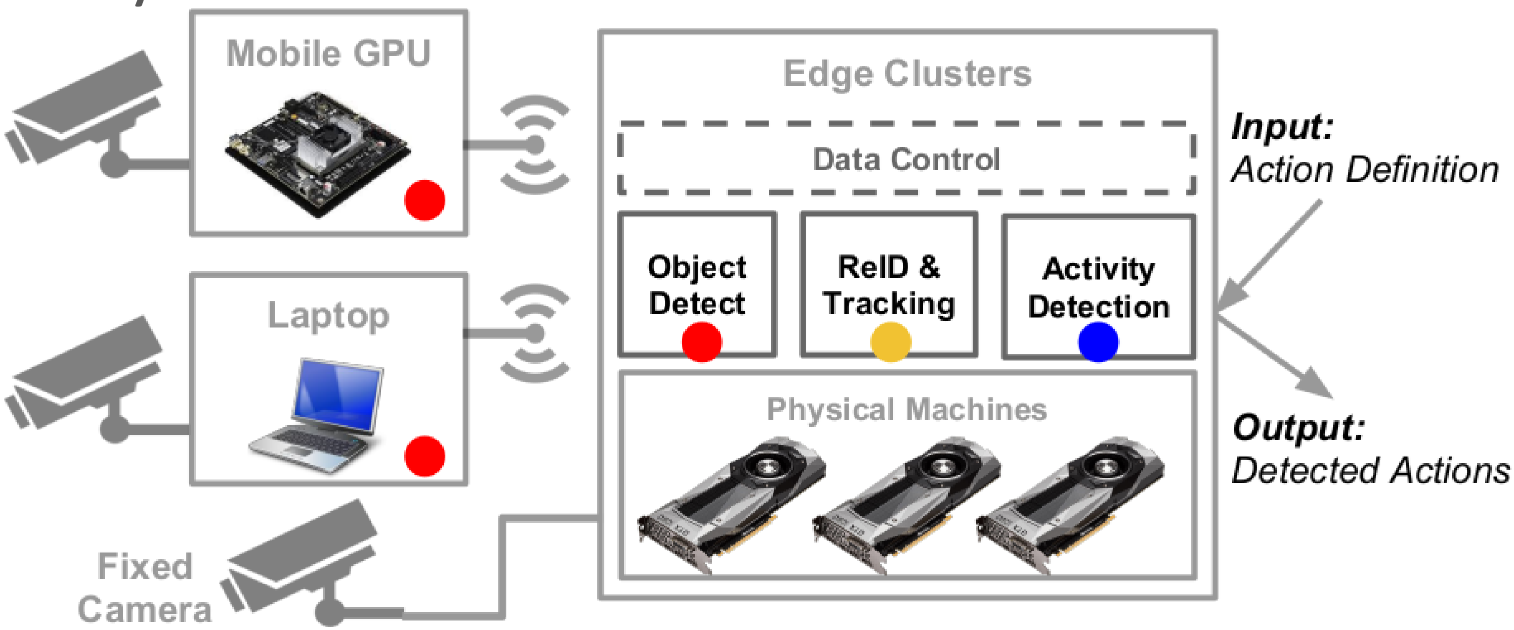
- Caesar: Cross-camera Complex Activity Recognition (Liu, 2019)
- Edge computing based system for complex activity detection.
- Provides vocabulary of activities to allow users to specify complex actions in terms of spatial and temporal relationships between actors, objects, and activities.

Vision-based HAR

- The output of Caesar

Vision-based HAR

- The system overview of Caesar.



	Input	Output
Object Detection	Image	Object Bounding Boxes
Track & ReID	Object Bounding Boxes Image	Object TrackID
Action Detection	Object Boxes & TrackID Image	Actions

Thank you!