Thesis for the Degree of Doctor of Philosophy August, 2013

Comprehensive Human Activity Recognition Framework based on Smartphone Multimodal Sensors

Manhyung Han

Advised by
Dr. Sungyoung Lee

경희대학교
KYUNG HEE UNIVERSITY

UCL Ubiquitous Computing Laboratory
Kyung Hee University, Korea

Contents

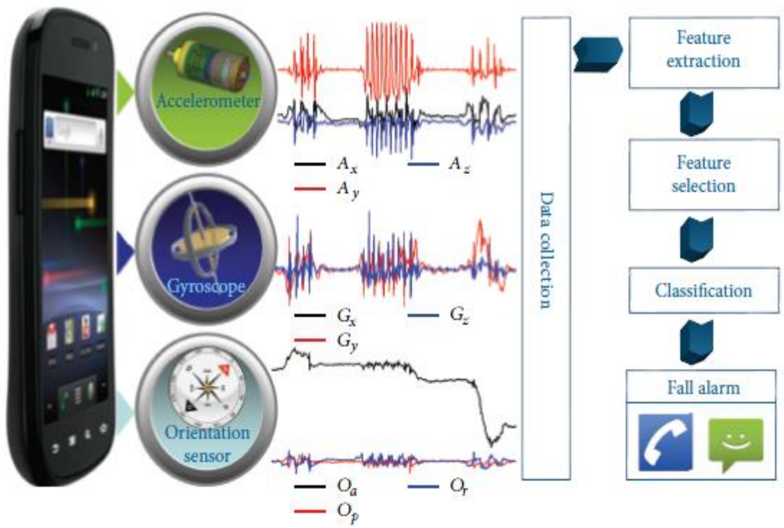
Motivation

- **Activity recognition using Smartphone multimodal sensors**
  - **AR with multimodal Sensors** – Utilizing multiple sensors (Accelerometer, Gyroscope, Audio, GPS, Wi-Fi, Light, Proximity etc.)
  - **Recognizing ADLs** – Physical movement, situation, context and locations
  - **Real-time Processing** – Real-time recognition on mobile devices, applicable to emergency services such as 'Fall detection'

- **Comprehensive AR framework**
  - **Novel framework** – Employ Accelerometer & Audio classifier for enhancing accuracy and recognizing various activities
  - **Utilizing multiple sensors** – Support both legacy recognition system and mobile environment - Smartphone
  - **Hierarchical approach** – For utilizing accelerometer and audio classification techniques in energy-aware system

Problem Statement

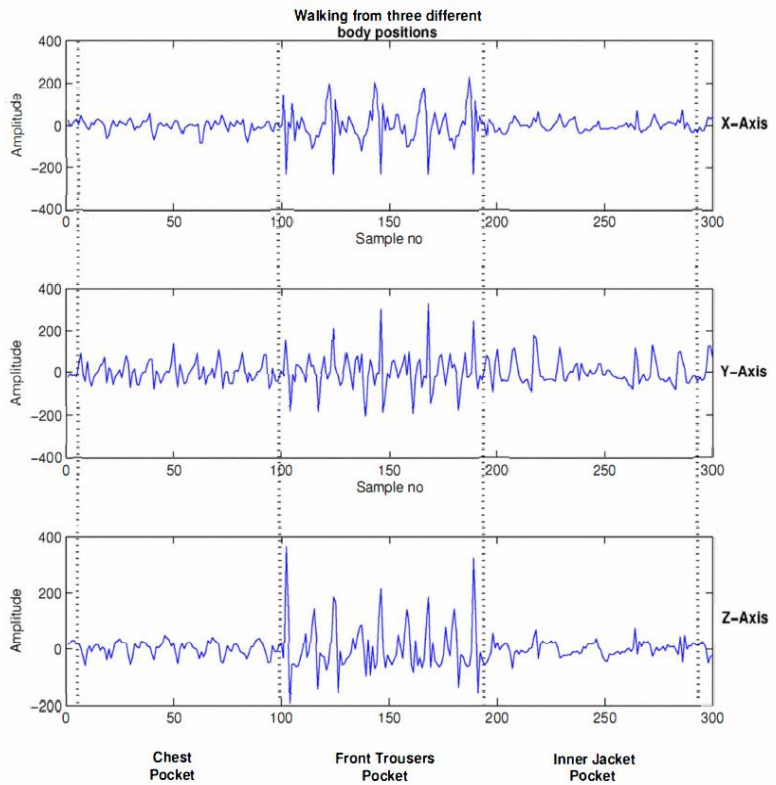
- **Considering single or simple sensor** – Most of the studies used single or simple sensor data(accelerometer, audio, video, GPS or light). <u>Hard to recognize various contexts.</u>[3],[16]

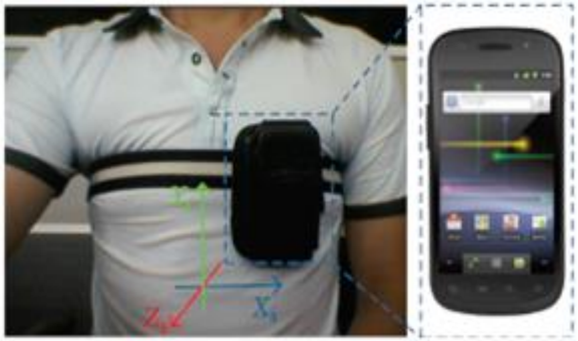


Recognizable activities: Yi He *et al*(2013)
Lying, sitting/standing, lie-to-sit, sit-to-lie, sit-to-stand
stand-to-sit, walking, walking upstairs, walking downstairs
Running, jumping, forward fall, right-side fall, backward fall, left-side fall

Problem Statement

- **Position-aware** – Concrete model for specific activity. Low flexibility and poor result in real world environment.[16],[17]





Position-dependent approach:
required to attach body-worn sensors or smartphone on specific location of the human body

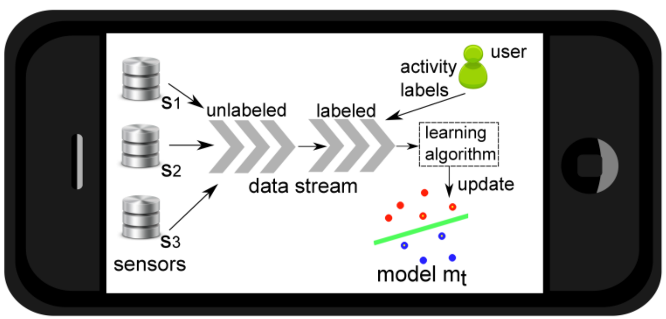
Problem Statement

- **Off-line processing**

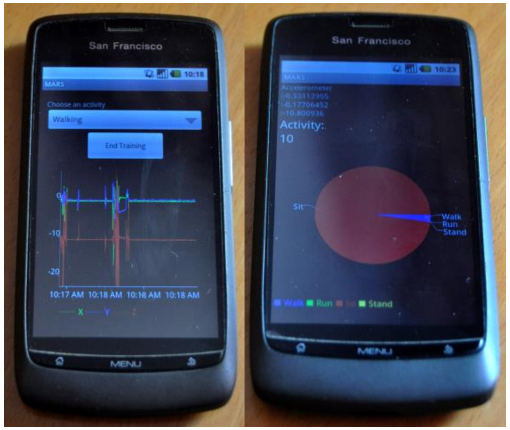
  ❖ Data gathering and processing / activity training and recognizing are separated.

  ❖ It may cause 'privacy problem' because the data is required to sent to an external site[13]

  ❖ At state-of-the-art researches, AR performing purely on a smartphone are initial stage(prototype, design) and poor accuracy [1],[14],[15]



Prototypes of Smartphone-based AR
activity recognition on smartphone is still initial stage and required to be enhanced for acceptable accuracy



Contributions

- **Utilizing multimodal sensors**
  - ❖ Movements, Poses, Actions, Situation, Context can be recognized.
  - ❖ Hierarchical approach for combining different sensor data classification results
  - ❖ Managing Accelerometer Classifier, Audio Classifier, Heuristic-based result validation
- **Novel framework for Smartphone**
  - ❖ Proposed a framework which utilizing multimodal sensors and recognizing activities in real-time.
  - ❖ The implementations of the framework support both legacy system and smartphone environment
- **On-line processing**
  - ❖ Activity and Context data is collected and processed on Smartphone. Real-time activity modeling & recognizing.

Related works

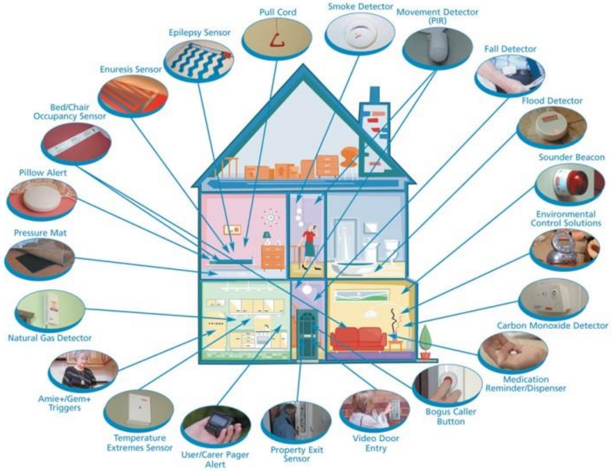
- **Sensors for Activity Recognition**

❖ Wearable sensor devices

- Generally, user is usually equipped with one or more devices(1 or more built-in sensors), placed at different parts of the body.
- Each device have processing, communication and storing ability.

❖ Environmental sensor devices

- Possible to recognize activities in a given place such as a room
- Sensors such as camera, microphone and RFID tags are deployed to provide implicit information that can be related to the activities



Related works

- **Sensors for Activity Recognition**

  - Combination of both wearable and environmental sensor devices

    - Sensor information from various type of sensors is aggregated and analyzed to produce activity and other useful contexts for potential further usages and adaptations.

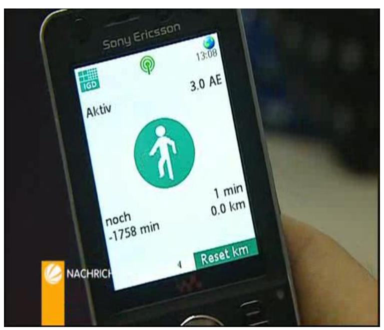
  - Multimodal sensors on smartphone

    - A kind of wearable sensors but various sensor devices are embedded on a single platform
    - Activity data is collected by multiple devices simultaneously.
    - Smartphone sensors does not required any dedicated sensor devices and the user always carrying it.

Smartphone with Sensors

Related works

- **Smartphone-based Activity Recognition**
  - ❖ Popular AR researches used several sensor devices commonly placed at different parts of the body
  - ❖ These sensors are placed at arm, wrist, knee, ankle and waist etc.
  - ❖ Investigations[3][52] have shown that accelerometer-based AR can give up to around 90% accuracy



Obtrusive for a person.
Users felt self conscious in public spaces and visually noticeable[7], [87], [105]

Related works

- **Smartphone as an alternative to current body-worn sensor**

❖ The available sensors are built-in(Flexibility and Readiness)

- Not required to use external sensors in order to collect activity information
- In cases where the need occurs, additional sensors and devices can be interfaced to smartphones

❖ Smartphones already have many properties that enable activity recognition related implementations.

- Relatively high processing power and sufficient memory for data processing
- Adequate storage space for the storage of raw and computed data
- Communication possibilities(Wireless network connection, Bluetooth etc.)

❖ Smartphone is likely to be with a user during daily activities

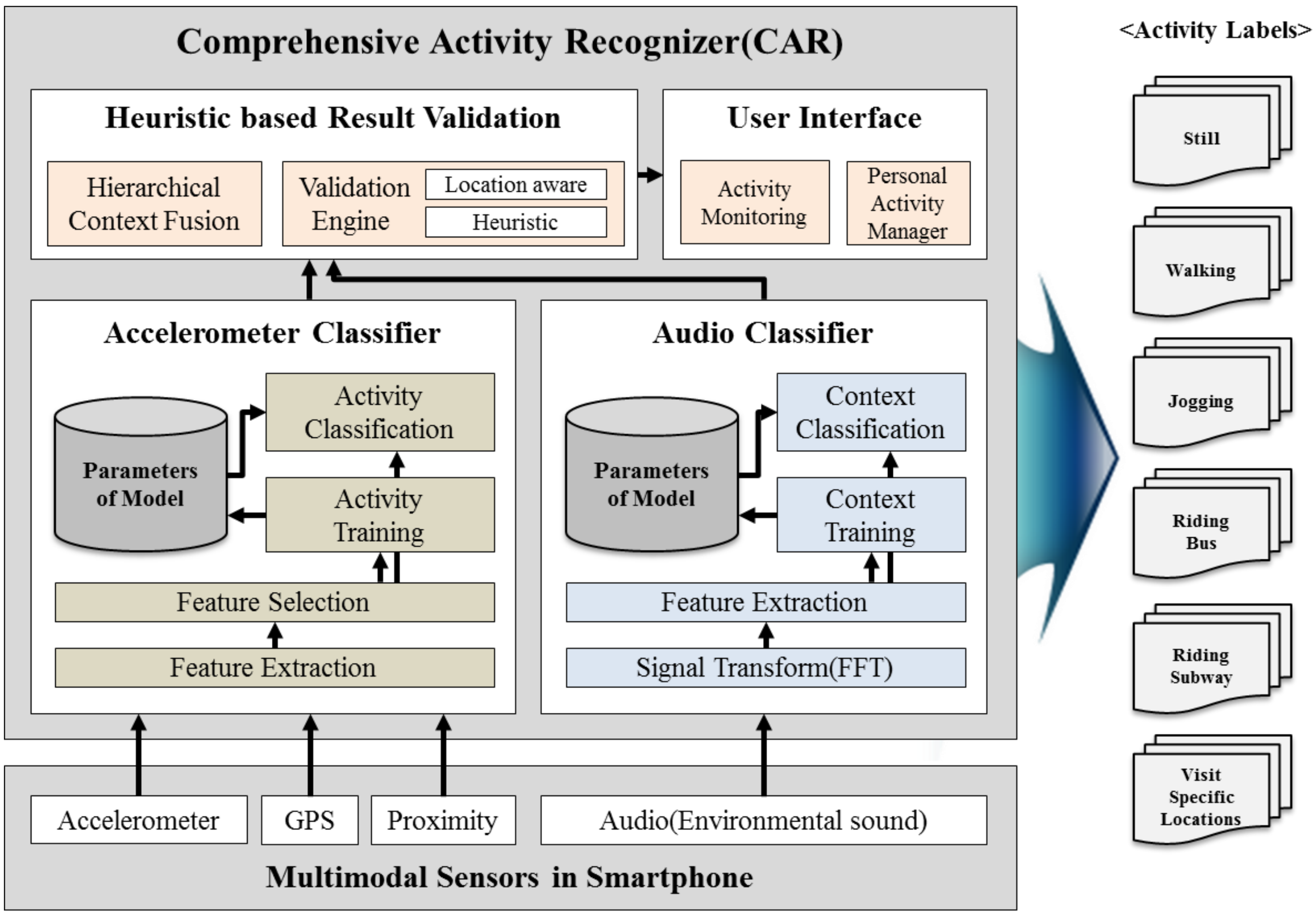
- Unobtrusive device

❖ Relatively long operation durations

- Whole day sensor data collection and processing is possible with proper management scheme

Comprehensive AR Framework

- **Proposed novel framework for utilizing smartphone multimodal sensors**



Comprehensive Activity Recognizer

- **Accelerometer data processing**

  ❖ Before modeling and classifying acceleration data, a prior process (feature extraction, selection) generates bunch of features

  ❖ A mixture model which is suitable for representing multiple distributions of collected data is chosen because of using multiple dimensions of features.

  ❖ Gaussian Mixture Model (GMM) is used for the acceleration data classification because it fits to process mean and variance value (position-free recognition)[113]

  ❖ Other classification techniques such as Gaussian Process are more appropriate for considering small number of variables or features.

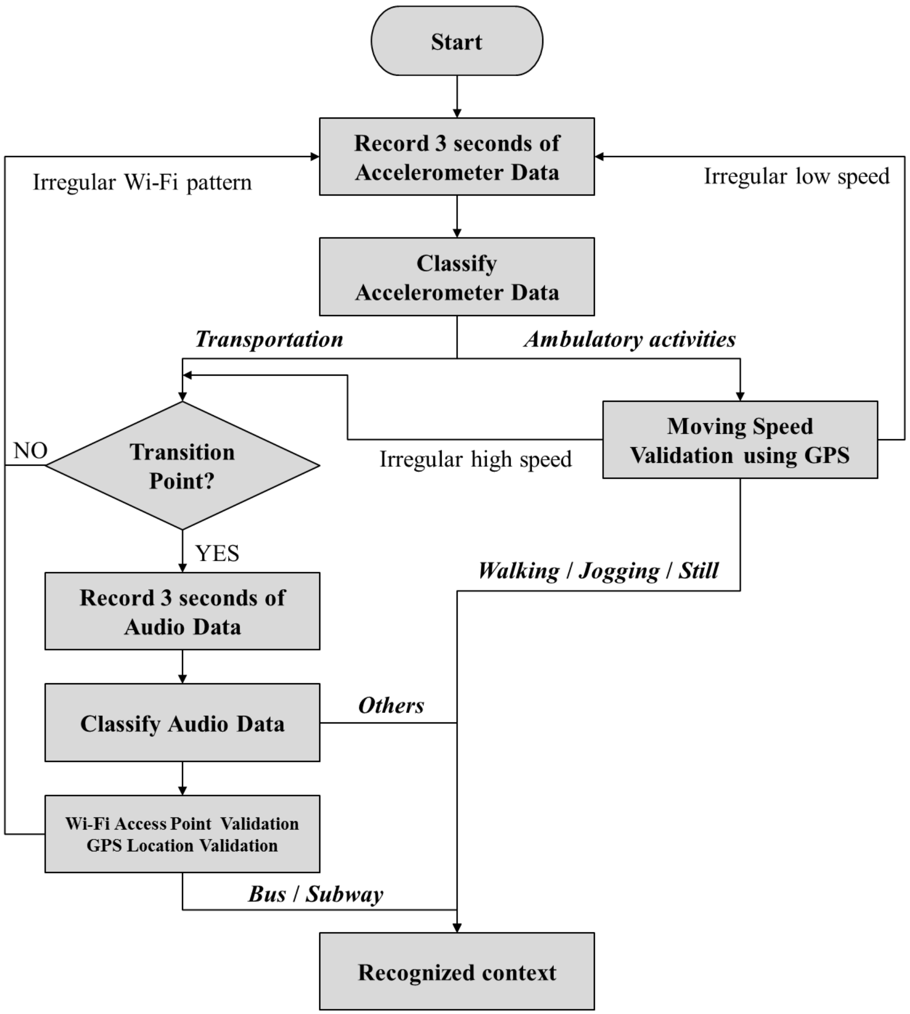
Comprehensive Activity Recognizer

- **Audio data processing**

  - For the audio classification, Hidden Markov Model (HMM) algorithm is used for training and testing audio data
  - Because the module needs to be classify only two activities which are bus and subway, and requires running on a smartphone in real-time[114].
  - There are other audio classification algorithms such as Conditional Random Field (CRF) and Support Vector Machine (SVM), but proposed approach using HMM is lighter than other algorithms
  - Also it fits in classifying similar audio data both collected from bus and subway

Comprehensive Activity Recognizer

- **Flow chart of the activity recognizer**



[Workflow]

- Classify accelerometer data as ambulatory activities(Walking, Jogging or still) or transportation using accelerometer classifier
- Then classify audio data as bus, subway or others using audio classifier
- Result validation by utilizing GPS and Wi-Fi data
- Result labels: Walking, Jogging, Still, Bus or Subway

Accelerometer Classification

- **Applying GMM for classification**

❖ After extracting and selecting features, utilizing Gaussian Mixture Model (GMM) to determine the parametric probability density function of each class -> $p(X^C/\lambda^C)$ where X^C is a training data matrix and λ^C is the parameter set

$$p(X^C|\lambda^C) = \sum_{i=1}^{M} \omega_i N(X^C|\mu_i, \Sigma_i)$$

$$N(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu_i)'\Sigma_i{}^{-1}(x-\mu_i)\right\}$$

❖ During the training phase, the parameters $\lambda^C = \{\omega, \mu, \Sigma\}$ are determined to maximize the training data likelihood $p(X^C|\lambda^C)$

❖ In the inference phase, given all the class parameter sets λ^{C1}, λ^{C2}, … , λ^{Cm} and an input vector x, the class label is determined by:

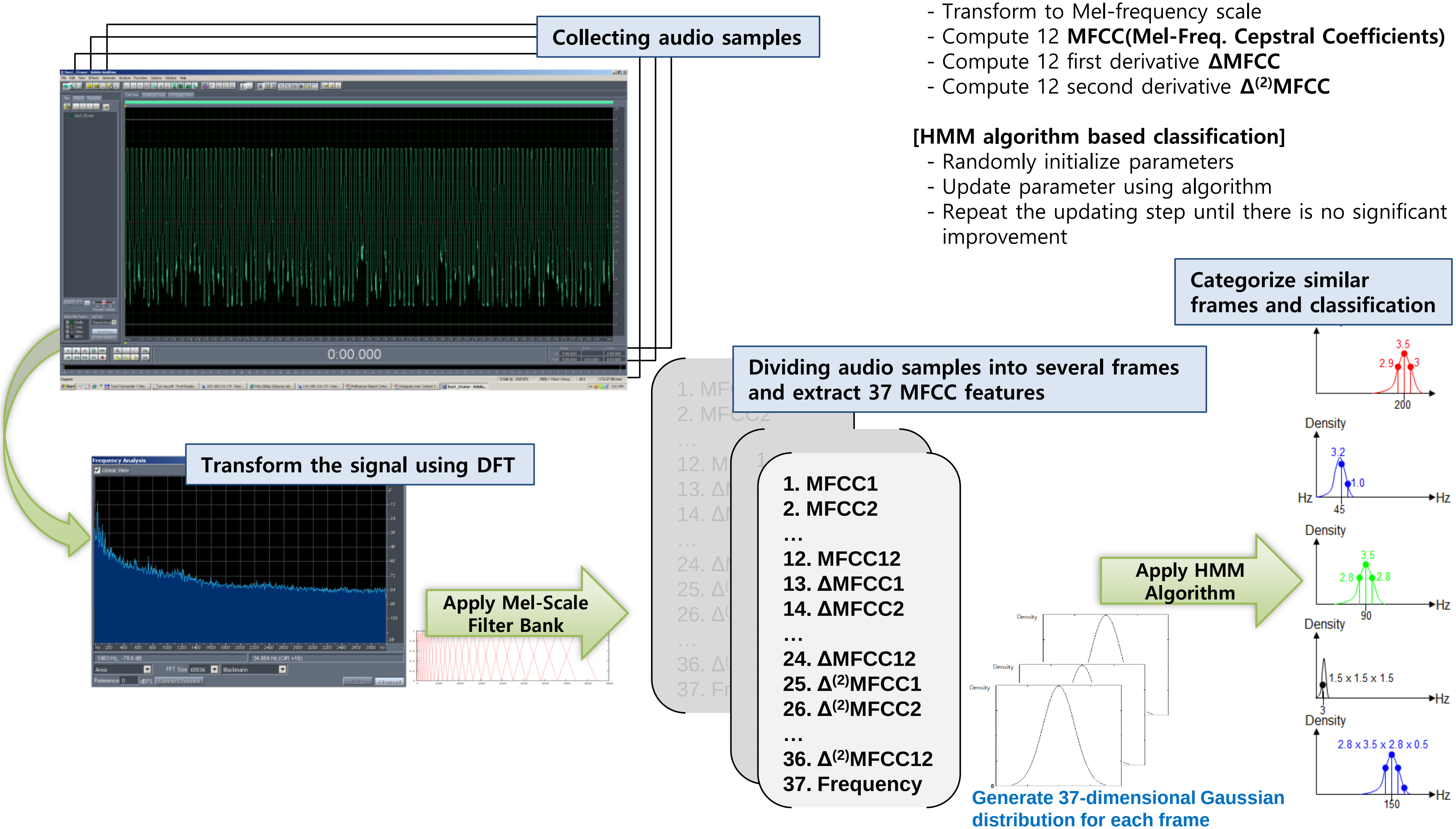
$$C = argmax_C(p(x|\lambda^C))$$

Audio Classification

- **Audio data classification workflow**

[Feature Extraction]
- Construct frequency-domain signal using Fast Fourier transform(**FFT**)
- Transform to Mel-frequency scale
- Compute 12 **MFCC(Mel-Freq. Cepstral Coefficients)**
- Compute 12 first derivative **ΔMFCC**
- Compute 12 second derivative **$\Delta^{(2)}$MFCC**

[HMM algorithm based classification]
- Randomly initialize parameters
- Update parameter using algorithm
- Repeat the updating step until there is no significant improvement



Audio Classification

- **Applying HMM for classification**

❖ HMM determines the characteristics of the data sequence. Parameter set is defined as

$$\Lambda = \{\pi, A, B\}$$

π = 1 ✕ N Vector of prior probability
A = N ✕ N transition probability matrix
B = set of N Observation density function

❖ And B is defined as

$$B(i,x) = \sum_{m=1}^{M} \omega_m G(x, \mu_m, \Sigma_m)$$

i = state index (1,2 … N)
M = No of Gaussian components
ω_m = mixing weight of the m[th] Gaussian component
$G(x, \mu_m, \Sigma_m)$ = Gaussian density function, mean and covariance matrix

❖ In the training phase, the model parameters are updated to maximize the training likelihood $P(X|\Lambda)$. After the training phase, each audio class has a corresponding HMM defined by the parameter sets $\Lambda^{Subway}, \Lambda^{Bus}, \Lambda^{Other}$

❖ In the inference phase, the likelihood of X can be computed by:

$$P(X|\Lambda^C) = \sum_{h_1,h_2,\dots,h_T} \pi(h_1)B(h_1,x_1)\prod_{t=2}^{T} A(h_{t-1},h_t)B(h_t,x_t)$$

h_t = hidden state value at time t

❖ So, final class label is decided by $C = Argmax_{C\in\{Bus,Subway,Other\}} P(X|\Lambda^C)$

Experimental result

- **Testing environment**
  - ❖ For testing the accelerometer classifier, over 10,000 data samples are collected from 10 volunteer graduate students during a month-long.
  - ❖ Position free data collection from smartphone multimodal sensors

Activity type	Sensor	Data format	No. of samples
Walking	Accelerometer, GPS, Wi-Fi	Text	1244
	Audio	Raw	1244
Jogging	Accelerometer, GPS, Wi-Fi	Text	591
	Audio	Raw	591
Bus	Accelerometer, GPS, Wi-Fi	Text	4645
	Audio	Raw	13023
Subway	Accelerometer, GPS, Wi-Fi	Text	3864
	Audio	Raw	3387
Car	Accelerometer, GPS, Wi-Fi	Text	955
	Audio	Raw	2829
Others	Accelerometer, GPS, Wi-Fi	Text	3106
	Audio	Raw	5472
Total number of sample(Accelerometer etc.)			**14405**
Total number of sample(Audio)			**26546**

Experimental result

- ## Audio Classification

  - To verify the audio classifier, 10 fold cross-validation has applied to the test dataset from University of East Anglia, UK
  - 8 kHz, 8 bit, mono WAV files taken using a Samsung YP55H MP3 recorder in 2004. Available at (http://lemur.cmp.uea.ac.uk/Research/noise_db/)
  - Also for evaluating of the classification for the activity 'transportation', own dataset from android smartphone is used.

(Avg. accuracy: 97.43%)

	Building Site	Bus	Car(City)	Supermarket	Office	Presentation	Street(Traffic)	Total
Building Site	100%	-	-	-	-	-	-	100%
Bus	-	100%	-	-	-	-	-	100%
Car	-	4%	95%	1%	-	-	-	100%
Supermarket	-	-	-	100%	-	-	-	100%
Office	-	-	-	-	100%	-	-	100%
Presentation	-	-	-	-	-	99%	1%	100%
Street	-	-	-	1%	1%	10%	88%	100%

(Avg. accuracy: 90.85%)

	Bus	Subway	Other
Bus	89.34%	5.60%	10.66%
Subway	4.25%	91.20%	4.55%
Other	4%	4%	92%

[Accuracy table using Ma, L. dataset and classification result for transportation activities using own dataset]

Experimental result

- **Comprehensive Activity Recognizer**

  - Both classifiers were combined into one integrated system – Comprehensive Activity Recognizer - with extra information acquired from the GPS and Wi-Fi

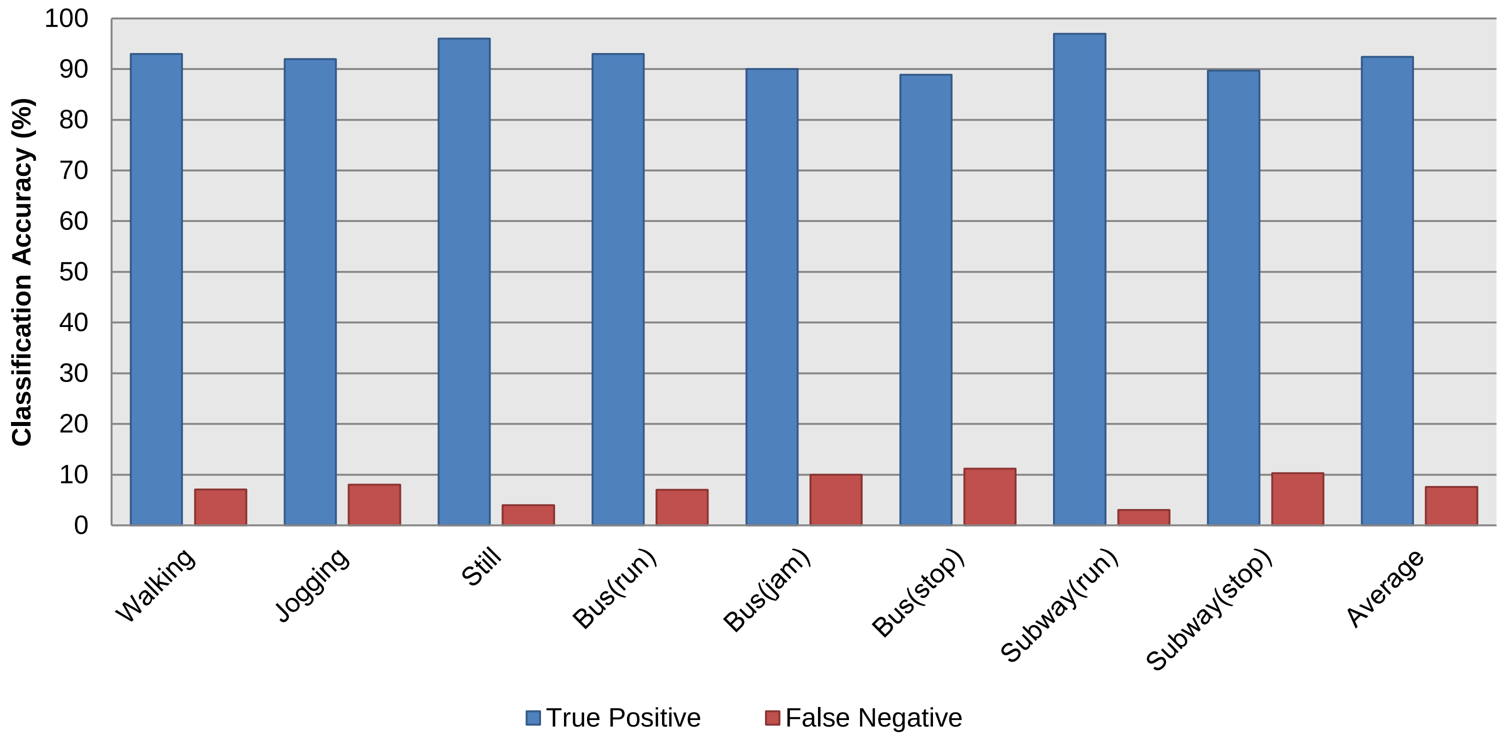
(Avg. accuracy: 92.43%)

	Ambulatory Activities			Bus			Subway		Total Samples
	Walk	Jogging	Still	Run	Jam	Stop	Run	Stop	
Walk	1109 (93%)	36	48	-	-	-	-	-	1193
Jogging	25	767 (92%)	42	-	-	-	-	-	834
Still	-	-	1915 (96%)	-	-	-	20	60	1995
Bus(run)	65	86	-	2000 (93%)	-	-	-	-	2151
Bus(jam)	-	-	52	-	782 (90%)	-	-	35	869
Bus(stop)	-	-	16	-	-	279 (89%)	-	19	314
Subway(run)	-	-	24	-	49	-	2341 (97%)	-	2414
Subway(stop)	-	-	18	-	11	7	-	314 (90%)	350

Experimental result

- **Comprehensive Activity Recognizer**

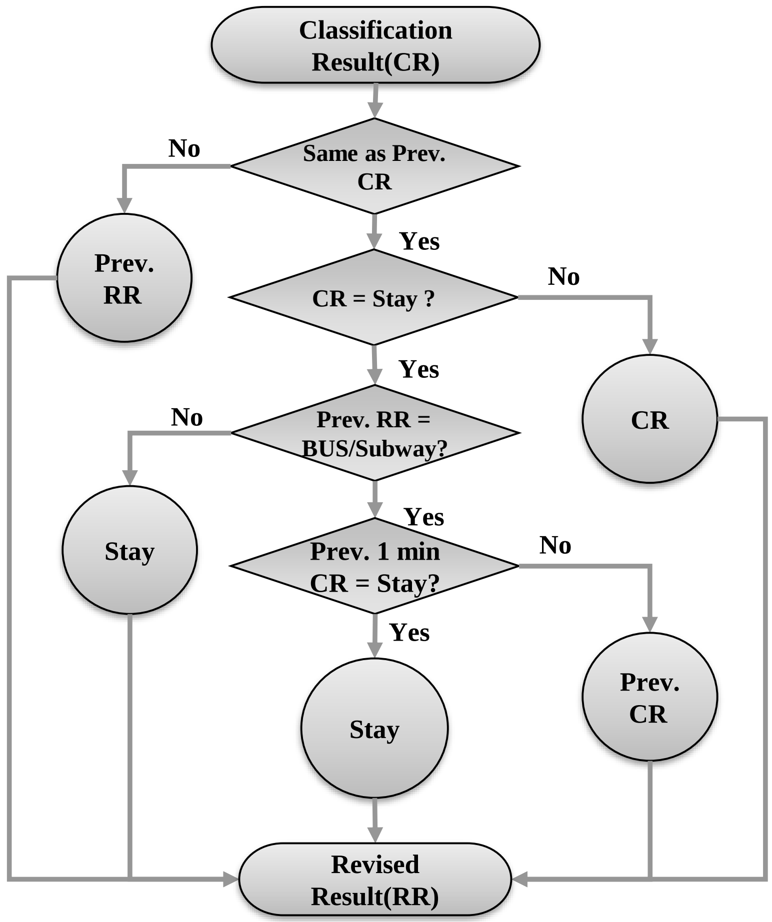
  ❖ Below is a comparison graph of the true positive with false negative of each activity. Average accuracy is 92.43%



Heuristic Approaches

- **Enhanced Decision Making and Revision Process**

❖ In real world environment, the accuracy is awfully low because of unexpected situations



[Description]

- To enhance the recognition accuracy, consider the expected situations and state changes
- CR(Classification Result): Label from comprehensive context recognizer
- RR(Revised Result): Revised Label using heuristic algorithm

Heuristic Approaches

- **Enhanced Decision Making and Revision Process**

Case 1	CR	RR
1	Stay	Initialized
2	Stay	Initialized
3	Stay	Initialized
4	Walking	Initialized
5	Walking	Walking
6	Stay	Walking
7	Walking	Walking
8	Walking	Walking
9	Walking	Walking
10	Walking	Walking
11	Walking	Walking
12	Stay	Walking
13	Stay	Stay
14	Stay	Stay
15	Walking	Stay
16	Walking	Walking
17	Bus	Walking
18	Bus	Bus
19	Bus	Bus
20	Bus	Bus
21	Bus	Bus
22	Stay	Bus
23	Stay	Bus
24	Bus	Bus
25	Bus	Bus
26	Bus	Bus
27	Stay	Bus
28	Stay	Bus
29	Bus	Bus
30	Bus	Bus

Case 2	CR	RR
…		Walking
133	Walking	Walking
134	Stay	Walking
135	Stay	Stay
136	Stay	Stay
137	Stay	Stay
138	Subway	Stay
139	Subway	Subway
140	Stay	Subway
141	Stay	Subway
142	Walking	Subway
143	Subway	Subway
144	Stay	Subway
145	Subway	Subway
146	Stay	Subway
147	Stay	Subway
148	Subway	Subway
149	Stay	Subway
150	Stay	Subway
151	Stay	Subway
152	Stay	Subway
…		Subway
167	Stay	Subway
168	Stay	Subway
169	Stay	Stay

**Comparison table of
Classification Result and Revise Results**

Heuristic Approaches

- **Revision results by heuristic approaches**

Before revision Case 1(up) / Case 2(down)

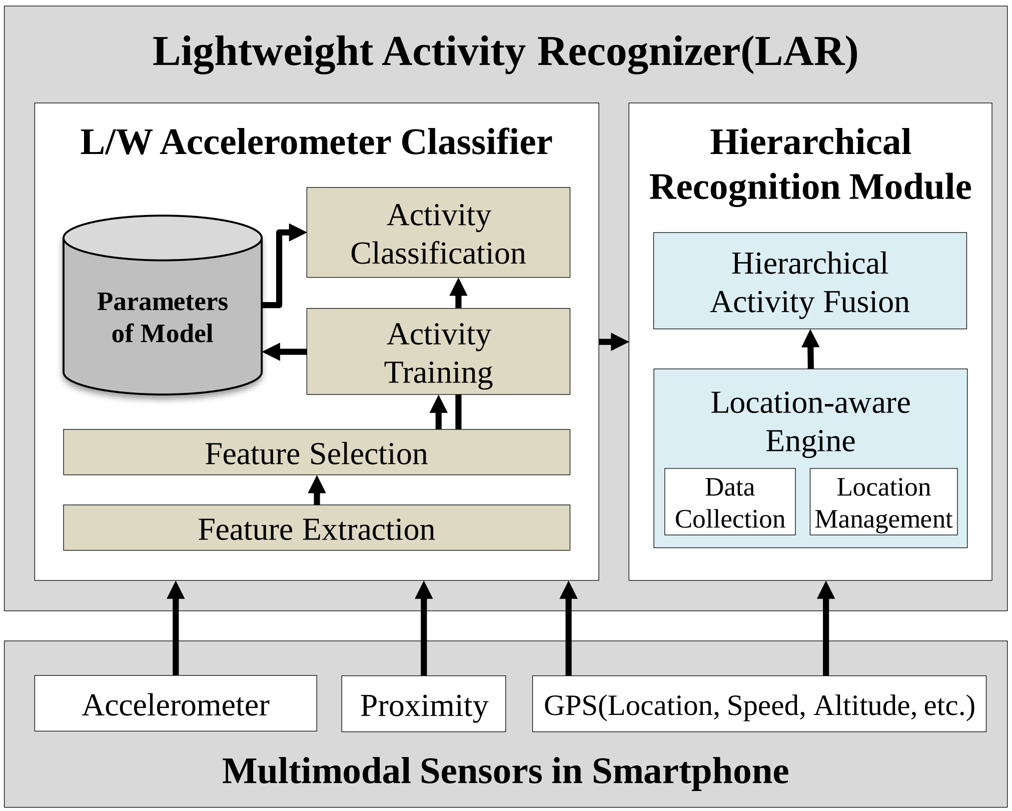
After revision Case 1(up) / Case 2(down)

Discussions

- Comprehensive AR Framework is proposed. It combines accelerometer classifier, audio classifier and Heuristic-based result validation module
- Accelerometer classifier using GMM classifies different ambulatory activities – Still, Walking, Jogging
- Audio classifier with MFCC and HMM is utilized for recognizing transportation activities using environmental sound
- For the evaluation, over 10,000 data samples are collected position freely from smartphone multimodal sensors
- Experimental results shows that HMM-base audio classification shows an accuracy of 90.85% in classifying bus and subway
- The accuracy of comprehensive activity recognizer is 92.43%

Lightweight AR Framework

- Activity modeling & recognition algorithm(A-NB) and hierarchical activity recognition framework(HARF) for a Smartphone environment
- Hierarchical approach for handling multimodal sensor data is proposed.



[Description]

- For running on resource restricted mobile device such as a smartphone, Audio data is not used
- As achieving high accuracy, hierarchical approach is proposed.
- The implementation provides personalized activity recognition by UI in smartphone

Lightweight Activity Recognizer

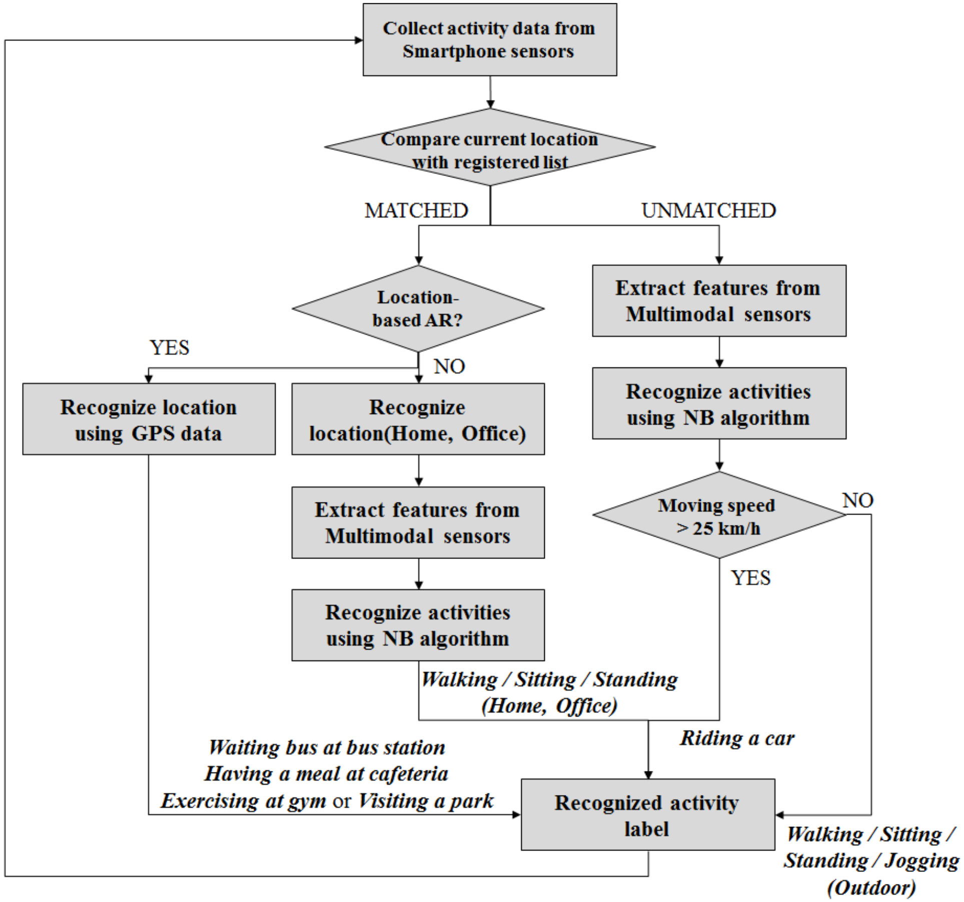
- **Proposed Framework(HARF)**
  - ❖ Another solution to resolve the problem 'the curse of dimensionality' when multiple sensor data are used-> Hierarchical classification framework is proposed
  - ❖ Enable to recognize complicated activities or contexts. One activity can be different up to the location information such as 'Sitting – Working – Resting – Riding a car'
  - ❖ Categorize activities into 3 groups
    - Location & Sensor based AR: ADLs in registered location
    - Location based AR: Visit specific locations(Life logs)
    - Heuristic based AR: Riding a car(GPS and heuristic rules re required)
  - ❖ Also considering heuristic methodology

Lightweight Activity Recognizer

- **Hierarchical Activity Recognition Framework(HARF)**

Type	Area	Activity	Sensors
Location & Multimodal sensor based activity recognition	Home	Walking	Accelerometer, Gyroscope, Proximity and GPS
		Sitting	
		Standing	
	Office	Walking	
		Sitting	
		Standing	
	Outdoor	Walking	
		Sitting	
		Standing	
		Jogging	
Location based activity recognition	Outdoor	Waiting bus at bus stop	GPS
		Having a meal at cafeteria	
		Exercising at gym	
		Visiting a park	
Heuristic based activity recognition	Outdoor	Riding a car	Accelerometer, Gyroscope, Proximity, GPS and Heuristic Rule

[Activity categorization for hierarchical activity recognition]



[Hierarchical activity recognition framework]

Lightweight Activity Recognizer

- **Adaptive Naïve Bayes algorithm(A-NB)**
  - ❖ Lightweight activity modeling & recognition algorithm based on Naïve Bayes
  - ❖ Executable on mobile device which has small memory space, cache but sufficient processing power -> Repetitive Calculation
  - ❖ Fast processing for real-time recognition

[Equations of Naïve Bayes]

$$p(C|F_1, \dots, F_N) = \frac{p(C)p(F_1, \dots, F_N|C)}{p(F_1, \dots, F_N)}$$

$$V_{max} = p(C_j|F_1, \dots, F_N) = p(C)\prod_{i=1}^{n} p(F_i|C)$$

$$P(F_i = v|C) = \frac{1}{\sqrt{2\pi\sigma^2{}_C}} e^{-\frac{(v-\mu_C)^2}{2\sigma^2{}_C}}$$

[Equations of A-NB]

$$\mu_N = \frac{\mu_{(N-1)} \times (N-1) + F_N}{N}$$ *Mean*

$$\mu'_N = \frac{\mu'_{(N-1)} \times (N-1) + F^2{}_N}{N}$$ *Mean Square*

$$\sigma^2{}_N = \mu'_N - \mu^2{}_N$$ *Variance*

$$V'_{max} = p(C_j|F_1, \dots, F_N) = p(C)\prod_{i=1}^{n} p'(F_i|C)$$

$$P'(F_i = v|C) = \frac{1}{\sqrt{2\pi\mu_v}} e^{-\frac{(v-\mu_m)^2}{2\mu_v}}$$

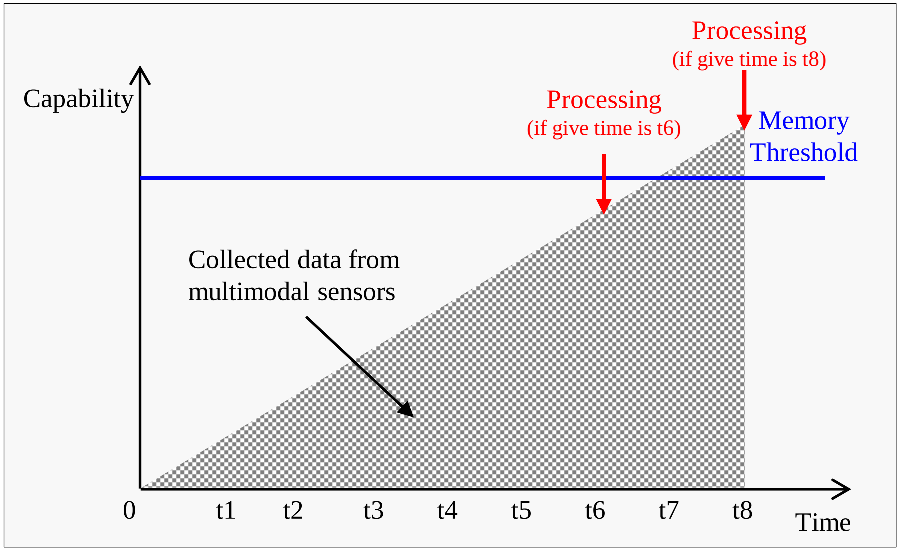
Mean of means

Mean of variances

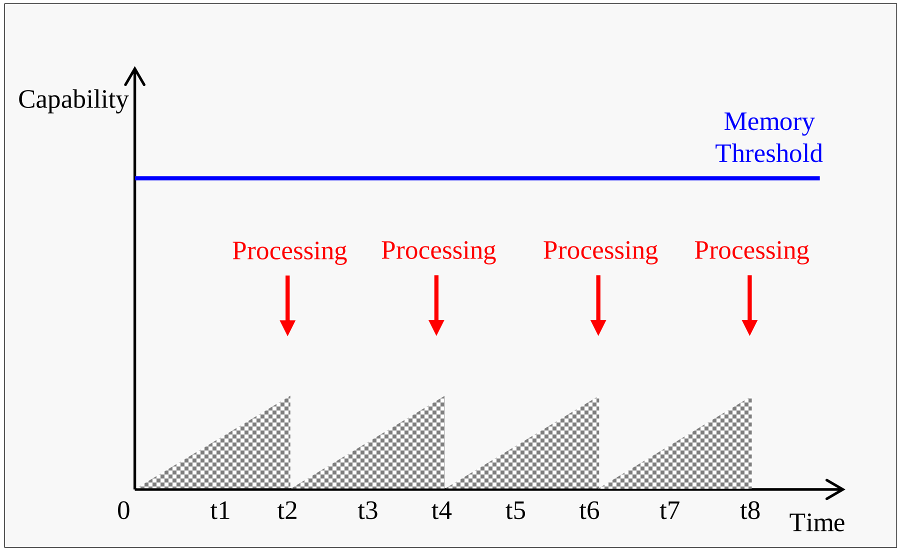
Lightweight Activity Recognizer

- **Adaptive Naïve Bayes algorithm(A-NB)**

❖ **Repetitive Calculation** – Periodically calculate accumulated sensor data for avoiding memory problem. But it caused frequent calculation. Lower the allocated memory, higher the processing frequency as figures below



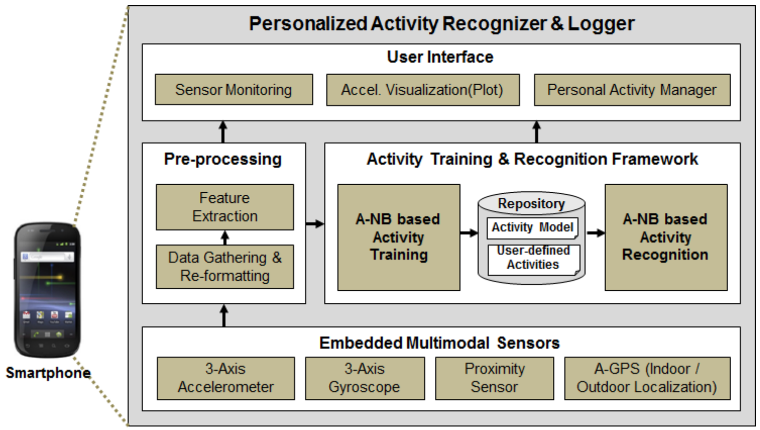
[Memory usage using Naïve Bayes algorithm]



[Memory usage and processing frequency using adaptive Naïve Bayes algorithm]

Implementation

- **Personalized Activity Recognizer & Logger(pARNL)**
  - ❖ Smartphone app. for testing and evaluating proposed algorithm(running on Android OS ver. 2.3.3, API level 10)
  - ❖ Provide on-line activity modeling interface for recognizing user's own activities or contexts
  - ❖ Implementing proposed HARF by utilizing multimodal sensors embedded in Smartphone

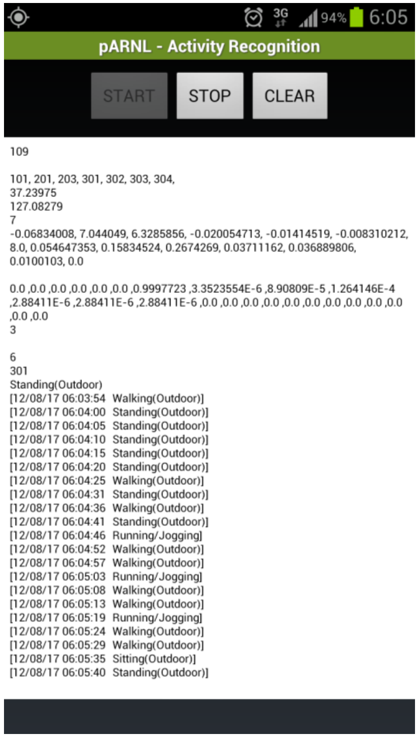


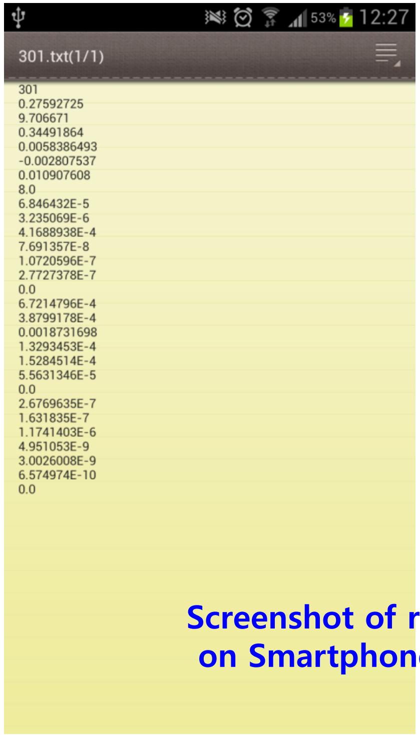
[HARF based real-time activity recognition system]

[Smartphone app. implementing real-time AR framework]

Experimental result

- **Evaluation setting for HARF**
  - ❖ Used features – Mean and Variance of 3-axis accelerometer, 3-axis gyroscope, GPS and proximity sensor
  - ❖ For testing data samples are collected from 3 volunteer graduate students during several weeks.





Screenshot of real-time recognition on Smartphone and data structure

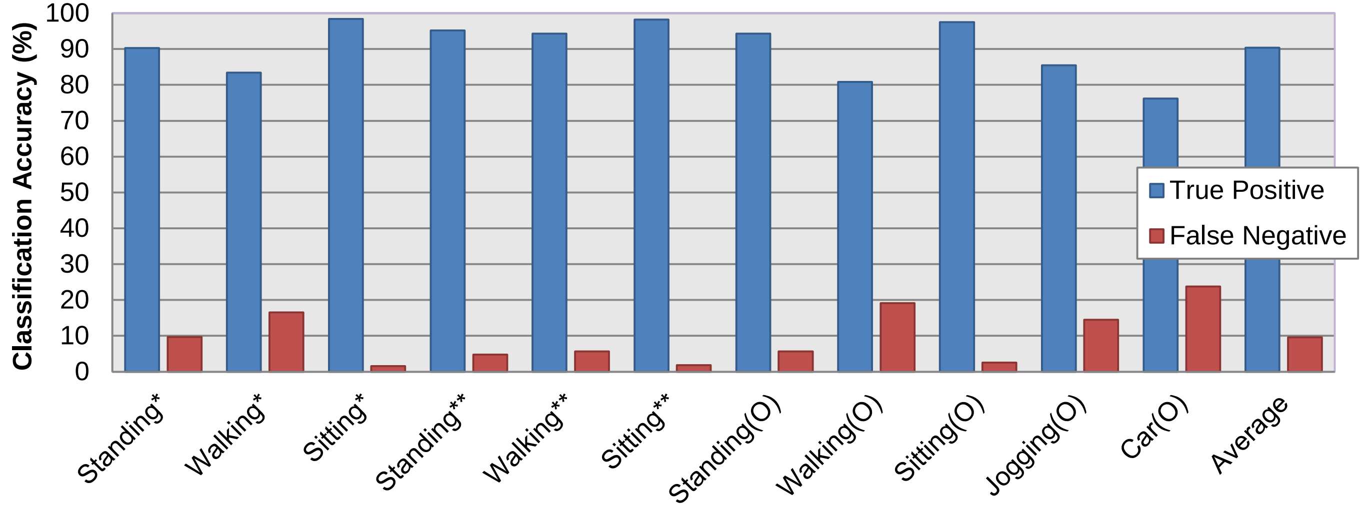
Experimental result

Avg. accuracy: 90.4%

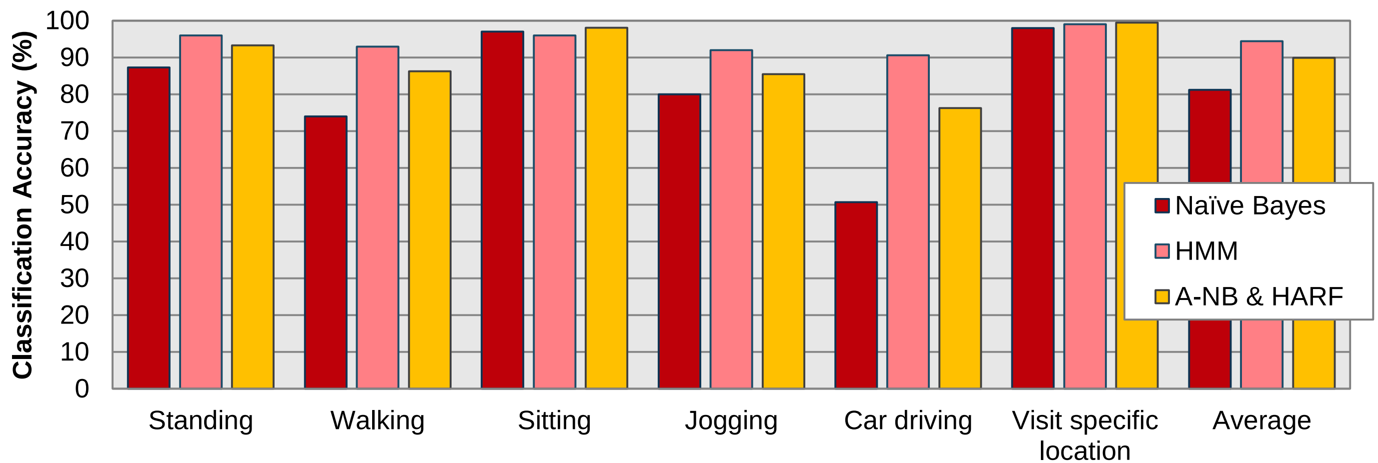
Location		Home			Office		Outdoor					
	Activity	Standing	Walking	Sitting	Standing	Walking	Sitting	Standing	Walking	Sitting	Jogging	Car
Home	Standing	**90.32**	-	9.68	-	-	-	-	-	-	-	-
	Walking	10.43	**83.47**	6.1	-	-	-	-	-	-	-	-
	Sitting	2.56	-	**98.44**	-	-	-	-	-	-	-	-
Office	Standing	-	-	-	**95.2**	-	4.8	-	-	-	-	-
	Walking	-	-	-	4.84	**94.35**	0.81	-	-	-	-	-
	Sitting	-	-	-	1.2	0.61	**98.19**	-	-	-	-	-
Outdoor	Standing	-	-	-	-	-	-	**94.34**	-	5.66	-	-
	Walking	-	-	-	-	-	-	12.77	**80.85**	6.38	-	-
	Sitting	-	-	-	-	-	-	2.5	-	**97.5**	-	-
	Jogging	-	-	-	-	-	-	2.17	10.86	1.47	**85.5**	-
	Car	-	-	-	-	-	-	16.25	6.25	1.25	-	**76.25**

[Activity recognition accuracy table of 11 activities for validating proposed HARF]

Experimental result



[Result 1. Activity recognition accuracy graph of 11 activities]



[Result 2. Accuracy comparison between Naïve Bayes and HARF for 15 activities]

Results analysis

- Result 1 shows accuracy(True Positive & False Negative) of recognized activities using proposed algorithm and framework

- Among the results, recognizing a context 'Riding a Car' shows the worst performance. Actual action is quite similar to 'Standing, Walking or even Sitting' – More heuristic approach is required!

- Result 2 is a comparison chart with Naïve Bayes and HMM the most-widely-used algorithm.

- Better than NB but little poorer than HMM. But considering the result of HMM was performed at server environment, HARF is sufficient to substitute for HMM.

Discussions

- Optimal algorithm and framework for Smartphone which has a limitation of resources are proposed. Accuracy ↑ / Complexity ↓
- Lightweight activity modeling & recognition algorithm (A-NB) based on Naïve Bayes / Hierarchical classification approach(HARF) which enables to recognize complicated activities or contexts
- For the evaluation, pARNL based on HARF using multimodal sensors in a Smartphone is implemented
- Experimental results shows that the proposed algorithm can classify 11 activities with an accuracy of 90.4%
- Compare to Naïve Bayes, HARF shows better performance over 10.73% and similar results to heavy & powerful algorithm(HMM)

Conclusion

- Comprehensive activity recognizer combining accelerometer classification, audio classification and GPS/Wi-Fi validation is proposed.
- Recognizing physical movement and transportation activities by combining accelerometer classifier and audio classifier.
- Heuristic approach for enhancing comprehensive activity recognizer is proposed.
- Lightweight AR framework for handling multimodal sensor data(except audio), is proposed for classifying various activities and contexts using Smartphone.
- Experimental results show that accuracy of both comprehensive activity recognizer(Off-line processing) and lightweight activity recognizer(On-line processing) are 92.43% and 90.4% respectively.

감사합니다!