HASC-IPSC: Indoor Pedestrian Sensing Corpus with a Balance of Gender and Age for Indoor Positioning and Floor-plan Generation Researches

Katsuhiko Kaji
Graduate School of Engineering,
Nagoya University
1 Furo-cho, Chikusa-ku, Nagoya,
Aichi, 4648603 JAPAN
kaji@nuee.nagoya-u.ac.jp

Ryoji Ban
Graduate School of Engineering,
Nagoya University
1 Furo-cho, Chikusa-ku, Nagoya,
Aichi, 4648603 JAPAN
van@ucl.nuee.nagoya-u.ac.jp

Hodaka Watanabe
Graduate School of Engineering,
Nagoya University
1 Furo-cho, Chikusa-ku, Nagoya,
Aichi, 4648603 JAPAN
hodaka@ucl.nuee.nagoya-u.ac.jp

Nobuo Kawaguchi
Graduate School of Engineering,
Nagoya University
1 Furo-cho, Chikusa-ku, Nagoya,
Aichi, 4648603 JAPAN
kawaguti@nagoya-u.jp

UbiComp'13 Adjunct, September 8–12, 2013, Zurich, Switzerland.

http://dx.doi.org/10.1145/2494091.2495981

Abstract
Up till now, the majority of researches related to location estimation and floor plan creation have used different kinds of data and there has simply been no technique to compare the relative advantages and disadvantages. We collected indoor pedestrian sensing data of 100 people with a balance of gender and age. The data is part of the HASC corpus, free to use for research purposes.

Author Keywords
Indoor pedestrian data; activity recognition; smartphone; sensor

ACM Classification Keywords
H.2.8. Spatial databases and GIS

General Terms
Experimentation

Introduction
Research related to indoor location estimation is a hot topic. In recent years, with the increased capabilities of smartphones, it has become possible to use a variety of location-based services. Outdoors, location information can be obtained via GPS, but the problem is that GPS signals cannot reach indoors. Various technologies such

as RFID, infrared, WiFi and IMES are being researched as ways to give absolute position indoors, but have not entered society as general purpose location estimation technology[2,3,6]. Research to increase the degree of accuracy also continues to be actively carried out. Dead reckoning, a method of estimating relative movement locus using a smartphone's in-built sensors, is being researched [7]. It is possible to estimate the route taken from a start point by detecting left and right turns, number of steps etc. However, errors are cumulative and so it is necessary to combine this method with an absolute location estimation method and perform error correction. Also, for dead reckoning to be used practically, it is necessary for there to be more flexibility with the smartphone's fix position. A dead reckoning method where the smartphone is held in one's hand/pocket is being investigated [4,9].

Research into the creation of building floor plans is also being actively carried out [1,8]. A floor plan is required as the basis to achieve a navigation or location-based service that can be used indoors. Here, floor plan means a geometric structure that indicates the possible ways that a person can move within that building. Building management will not always release this structural information. Furthermore, even if the information is released, the cost of creating a network, generally from CAD data or floor map held by the building manager, is a problem. A method is proposed where a person with a smartphone moves around inside a building and a floor plan is created automatically based on that data.

Further advances in the two research topics of location estimation and floor plan creation are needed in order to provide various indoor location-based services for real world use. Up till now however, various methods for different environments have been proposed, but the majority use different data and there has simply been no technique to compare the relative advantages and disadvantages. The HASC corpus already contains real world activity data on the indoor and outdoor areas of stations, shops and offices [5]. However, this data does not give completely correct routes and there is not much behavioral data within each area. Therefore, it cannot be used as benchmark data for location estimation or floor plan creation.

We collected pedestrian data within buildings as the basis for our research. The data, named HASC-IPSC, is part of the HASC corpus, free to use for research purposes. You can download the corpus via "http://hub.hasc.jp". Here I will discuss the details of HASC-IPSC.

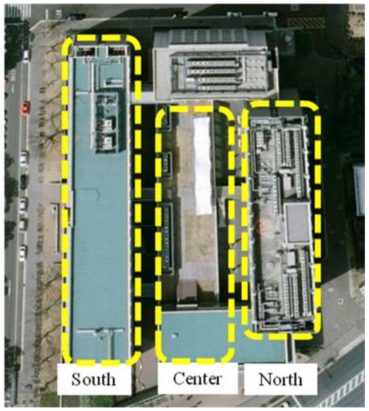

Figure 1. Target buildings for data collection.

Age	Male	Female
Early 20s	5	5
Late 20s	5	5
Early 30s	5	5
Late 30s	5	5
Early 40s	5	5
Late 40s	5	5
Early 50s	5	5
Late 50s	5	5
Early 60s	5	5
Late 60s	5	5
Total		100

Table 1. Gender and age of subjects.

Indoor Pedestrian Moving Data

Our aim is to create common data for indoor location estimation, floor-plan generation, and activity recognition. Because age and gender affect activity sensor data, we intended to correct balanced data with a valance of age and gender.

We looked at the East Building (B1F-5F), the Central Building (B1F-2F) and the South Building (1F-3F) of the Nagoya University IB Integrated Building (Figure 1). The buildings are all adjacent to each other. In order to collect data with a balance of gender and age, we had subjects perform as Table 1.

We set the start and end points of route to be within the rooms or terminals of corridor of these buildings. We had each subject perform a total of four kinds of route, for example a round trip within one building and a round trip to and from a different building. We also had four different subjects move along each route.

When moving indoors, we confirmed the start and end points, and route with the subjects. We also had the subjects observe the following points while moving.

1. Try to walk normally, no running.
2. Try to stay in the center of corridors.
3. Open doors by yourself.
4. Do not use elevators or escalators, take the stairs.
5. Do not jump up or down the stairs, climb up one step at a time.

Devices used were the Nexus 4 (Android 4.2.1) and Galaxy S3 (Android 4.1.2). The Nexus 4 was placed in a holder on the back of the waist. The Galaxy S3 was carried as normal (e.g. in a shirt pocket or in a bag). The sensor information during the route was logged by the Android version of the HASC Logger installed on each device.

The sensor types and actual sampling rates of each smartphone can be seen in the Table 2.

	Nexus 4	Garaxy S3
Accelerometer	99Hz	88Hz
Gyro	99Hz	121Hz
Magnetics	99Hz	88Hz
Air pressure	99Hz	65Hz
Light	99Hz	293Hz
Proximity	99Hz	98Hz
WiFi	1.0Hz	1.8Hz

Table 2. Sensors and their sampling rates.

The collected data was labeled. When labeling, we input the route type and distance between each node, where a node is a left or right turn, passing through a door etc.

Activity type	walking, running, ascending stairs, descending stairs, skipping, standing still
Data type	**Segmented data** (mostly for training data): For each subject, 5 series of 20sec of activities.
	Sequence data (mostly for test data): For each subject, 120sec of labeled activity sequence which includes all of activities. (Each activity should longer than 5 sec).

Table 3. Brief summary of HASC corpus.

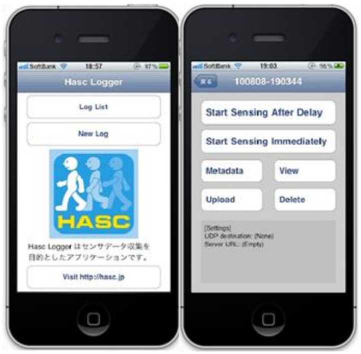

Figure 3. HASC Logger is available for free via App Store and Google Play.

A second person accompanied each subject, following behind to guide them and ensure they took the correct route. The accompanying person also pushed a button whenever a turn was made, a door was opened and when the destination was reached. The button pushes and navigation voice were recorded on the handheld HASC Logger (Figure 3) terminal and used afterwards in the labeling process. In order to synchronize the terminals of the subject and the accompanying person, when starting the route, the accompanying person called out and pushed the button on their terminal while the subject tapped their own terminal. When labeling, by matching the acceleration waveform from when the terminal was tapped and the timing of the button being pressed, the terminals of the subject and accompanying person could be synchronized.

As with the existing HASC corpus, we collected data on six kinds of basic activity (walking, running, ascending stairs, descending stairs, skipping, standing still) from the subjects. It can be used as training data for activity recognition. Table 3 shows brief summary of basic activity data in HASC corpus.

Formats

Every node such as terminal, door and turning points of corridor has ID. NodeID format is as shown below.

NodeID: [BuildingID]-[FloorID]-[NodeType]

Each NodeID has a number for the BuildingID (e.g. b1, b2). The FloorID is shown as f01, f02 etc. where the number denotes the floor number. The stair landing which connects Floor 1 and Floor 2 is shown as "f01f02." The NodeType follows the four English letters below, and each type is assigned a kind of unique serial number within that floor (e.g. t01, c01). For example, "b1-f01-c03" indicates the first building, first floor, third corridor node.

- t: terminal (start/end points of route)
- d: door
- c: corridor
- s: stair landing

```
b1-f01-c01,0.00,0.00,0.00,b1-f01-t01,6.10,0.00,0.00
b1-f01-c01,0.00,0.00,0.00,b1-f01-d02,0.00,-26.90,0.00
b1-f01-d02,0.00,-26.90,0.00,b1-f01-d04,0.00,-37.60,0.00
b1-f01-c05,0.00,-60.20,0.00,b1-f01-d04,0.00,-37.60,0.00
b1-f01-c05,0.00,-60.20,0.00,b1-f01-t04,6.70,-60.20,0.00
b1-f01-c03,0.00,-30.10,0.00,b1-f01-d04,0.00,-37.60,0.00
b1-f01-c03,0.00,-30.10,0.00,b1-f01f02-s02,7.97,-30.10,2.16
b1-f01f02-s01,7.97,-28.40,2.16,b1-f01f02-s02,7.97,-30.10,2.16
b1-f01f02-s01,7.97,-28.40,2.16,b1-f02-c01,3.30,-28.40,4.50
```

Figure 4. A part of coordinate .csv.

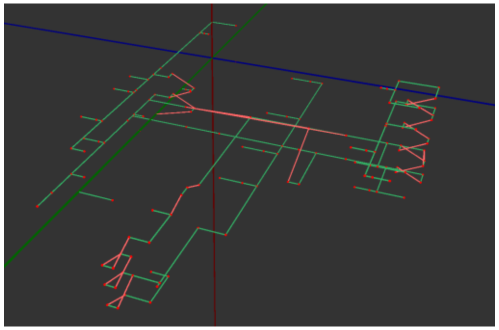

Figure 5. Whole floor plan of the target environment.

The 3D coordinate data from each NodeID is described in "coordinate.csv" (Figure 4). Each row expresses one link. Coordinates of two nodes of the link is described.

The whole structure of the environment is shown in Figure 5.

```
b1-f01-c05,b1-f01-d04,22.6
b1-f01-c05,b1-f01-t04,6.7
b1-f01-c03,b1-f01f02-s02,7.97,2.16,12
b1-f01f02-s01,b1-f01f02-s02,1.7
b1-f01f02-s01,b1-f02-c01,4.67,2.34,13
```

Figure 6. A part of linkinfo.csv

```
#targetfile:hasc-20130302-142434-acc.csv
#audiofile:hasc-20130302-142434-audio.wav;offsettime:0.0
1.36220207717E9,,b3-f02-t01;start
1.36220207717E9,1.362202083203E9,walk;4.7
1.362202083203E9,,b3-f02-c04;right
1.362202083203E9,1.362202089942E9,walk;8.4
1.362202089942E9,,b3-f02-c05;left
1.362202089942E9,1.362202096692E9,walk;7.5
1.362202096692E9,,b3-f02-d02;
1.362202096692E9,1.362202103533E9,walk;7.1
1.362202103533E9,,b2-f02-c07;left
1.362202103533E9,1.362202119847E9,stDown;12.2
1.362202119847E9,,b2-f01-c12;left
1.362202119847E9,1.362202121504E9,walk;2.0
1.362202121504E9,,b2-f01-c14;left
```

Figure 7. A part of label file.

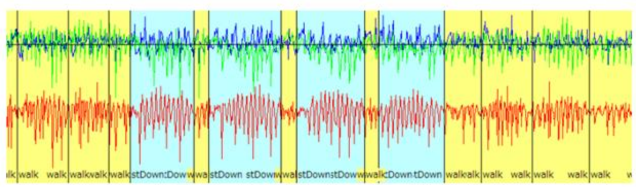

Figure 8. An example of labeled route data.

The link type can be identified from the NodeID. For example, if the FloorID differs between two NodeIDs that are connected by a link, then the link type is stairs. If the NodeType starts with 'd', then that link involves passing through a door.

The information of each link between nodes is described in linkinfo.csv (Figure 6). Every line contains two-dimensional length of the link. If the link involves stairs, the height and number of stairs is also listed.

An example of correct labeling assigned to route data is shown in Figure 7. Each row shows an event node or a link. If a row has no second component, then it is an event node. The third component is meta-information. In the case of a node, information on how the node was passed through is entered (start, right, left, end). If nothing is entered, then the node was passed straight through. In the case of a link, the link's activity type and distance are separated by a semicolon. By using HASC Tool, label information can be visualized as shown in Figure 8. According to label information, we can generate trajectory of each movement data (Figure 9).

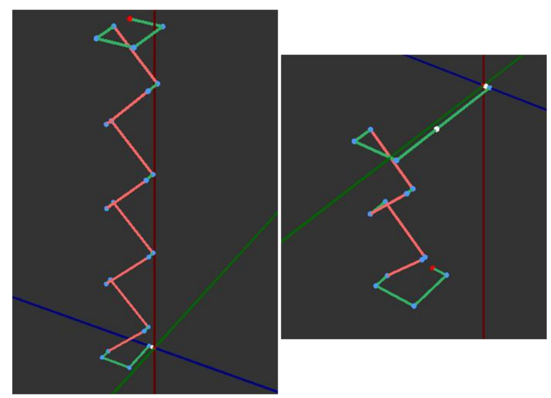

Figure 9. Trajectories of route data.

The subjects' personal data is described in the file "personal.meta". This includes information on age, gender, height, weight, number of items of clothing generally worn and shoe type.

# of Subjects	100
# of Routes	100
# of Terminals	32
# of Nodes per one route	Avg: 17.53 SD: 7.33
Time per one route	Avg: 109.44s SD: 40.15s
Length per one route	Avg: 82.56m SD: 30.72m
Time per one link	Avg: 6.62s SD: 4.89s
Length per one link	Avg: 5.00m SD: 5.35m
# of WiFi access points per one observation	Avg: 33.73 SD: 3.38

Table 4. Statistics of HASC-IPSC.

Statistics

Table 4 shows the statistics of HASC-IPSC. As you can see, we collected indoor pedestrian moving data on the scale of one hundred subjects.

Conclusion

This paper describes about indoor movement data that can be used as benchmark data for indoor location estimation and floor plan creation. The data was taken from 100 people with a balance of gender and age. It is used free of charge for research purposes. We would be glad if this data could be used in basic and applied research and contribute further to advancement in the fields of positional information processing and activity recognition.

Acknowledgement

A part of this work was supported by JSPS KAKENHI Grant Number 23240014.

References

[1] Alzantot, M., Youssef, M., CrowdInside: Automatic Construction of Indoor Floorplans, Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL'12), pp.99-108 (2012)

[2] Bahl, P., and Padmanabhan, V. N.: RADAR: An In-Building RF-based User Location and Tracking System, Proceedings of IEEE Infocom 2000, pp.775–784 (2000)

[3] Kaji, K., Kawaguchi, N., Design and Implementation of WiFi Indoor Localization based on Gaussian Mixture Model and Particle Filter, in Proceedings of The 3rd International Conference on Indoor Positioning and Indoor Navigation (IPIN2012), pp.1-9 (2012)

[4] Kamisaka, D., Muramatsu, S., Iwamoto, K., et al., Dead Reckoning Method by Hand for Pedestrian Navigation System, IPSJ Journal, Vol.52, No.2, pp.558-570 (2011)

[5] Kawaguchi, N, Yang, Y., Yang, T., et al., HASC2011corpus: Towards the Common Ground of Human Activity Recognition, in Proceedings of 13th ACM International Conference on Ubiquitous Computing, pp.571-572 (2011)

[6] Manandhar, D., Kawaguchi, S., Uchida, M., et al.: IMES for Mobile Users Social Implementation and Experiments based on Existing Cellular Phones for Seamless Positioning, International Symposium on GPS/GNSS (2008)

[7] Perry, J., Gait Analysis: Normal and Pathological Function, SLACK Incorporated, Thorofare (1992)

[8] Shin, H., Chon, Y., Cha, H. Unsupervised Construction of an Indoor Floor Plan Using a Smartphone. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 42, Issue.6, pp. 889–898 (2012)

[9] Steinhoff, U. and Schiele, B.: Dead reckoning from the pocket - An experimental study, Proceedings of Eighth Annual IEEE International Conference on Pervasive Computing and Communications, pp. 162–170 (2010)