HASC2011corpus: Towards the Common Ground of Human Activity Recognition

Nobuo Kawaguchi
Ying Yang
Tianhui Yang
Nobuhiro Ogawa
Yohei Iwasaki
Katsuhiko Kaji
Nagoya University
{kawaguti,yang,tenki,hiro,iwasaki,kaji}@ucl.nuee.nagoya-u.ac.jp

Tsutomu Terada
Kazuya Murao
Kobe University
tsutomu@eedept.kobe-u.ac.jp,
murao@port.kobe-u.ac.jp

Sozo Inoue
Kyushu Institute of Technology
sozo@mns.kyutech.ac.jp

Yasuyuki Sumi
Future University Hakodate
sumi@acm.org

Yoshihiro Kawahara
The University of Tokyo
kawahara@akg.t.u-tokyo.ac.jp

Nobuhiko Nishio
Ritsumeikan University
nishio@cs.ritsumei.ac.jp

ABSTRACT
Human activity recognition through the wearable sensor will enable a next-generation human-oriented ubiquitous computing. However, most of research on human activity recognition so far is based on small number of subjects, and non-public data. To overcome the situation, we have gathered 4897 accelerometer data with 116 subjects and compose them as HASC2011corpus. In the field of pattern recognition, it is very important to evaluate and to improve the recognition methods by using the same dataset as a common ground. We make the HASC2011corpus into public for the research community to use it as a common ground of the Human Activity Recognition. We also show several facts and results of obtained from the corpus.

Author Keywords Activity Recognition, Activity Understandings, Wearable Computing, Accelerometer, Wearable Sensor, Large Scale Corpus, HASC.

ACM Classification Keywords H.5.2 [User Interfaces]: Interaction styles (e.g., commands, menus, forms, direct manipulation)

General Terms Measurement, Experimentation, Human Factors, Documentation, Standardization

INTRODUCTION
Most of research on human activity recognition so far is based on small number of subjects and lab-created private data. So it is difficult to compare the methods/algorithms across the literatures. To overcome the situation, we have started a project named "HASC Challenge[1]" to collect a large scale human activity corpus. By the collaboration of more than 24 teams, we have gathered 4897 carefully and precisely labeled accelerometer data with 116 subjects and compose them as HASC2011corpus. We make the HASC2011corpus into public[2] for the research community to use it as a common ground of the Human Activity Recognition. In the following, we will explain the HASC[1] Activity data format which enables simple and easy data exchange. Then we show the basic information of the HASC2011corpus. We also show HASC Tool and the current preliminary results from the HASC2011corpus.

HASC ACITIVITY DATA FORMAT
To share the activity data or processing functions among the researchers and developers, activity data format must be standardized. We have defined the following data format as HASC data format for activity understandings.

Sensor data (.csv)
We defined sensor data file format as a simple csv format with time stamp and sensor values. For the accelerometer data, it may contain: time stamp, x, y and z axis-acceleration values for each row. Time stamp is in seconds with floating point. So any sampling rate data can be stored with this format. Accelerations are in the gravitational acceleration unit ($1G = 9.80665 m/s^2$).

Meta information format (.meta)
For each sensor data, related information of the subject and the data acquisition condition are important. We defined a meta information file format to record subject's gender, weight and height, and sensor's type, sampling rate and position. The style of the format is simple "attribute:value" pair.

Label data format (.label)
For each continuous activity data, "tag/label" is required to put on the activity time period. We defined a .label data format as a csv format with start-time, end-time and label-name. By using this format, one can easily add any kind of label onto the time-series data. However, definition of the labeling is not easy. We need further research on this area.

HASC2011CORPUS
HASC2011corpus contains following files.

UbiComp'11, September 17–21, 2011, Beijing, China.
ACM 978-1-60558-843-8/10/09.

[1] HASC: Human Activity Sensing Consortium.

Gender	Number of Subjects	Number of Files
Male	102	4464
Female	14	434
Total	116	4898

Table 1. Statistics of HASC2011corpus.

Segmented data (mostly for training data): For each subject, 5 series of 20sec of 6 activities: stay, walk, jogging , skip, stair-up(stUp), stair-down(stDown).

Sequence data (mostly for test data): For each subject, 120sec of labeled activity sequence which includes all of above 6 activities. (Each activity should longer than 5 sec).

So for each subject, we have at least 31 activity data files. (Some subjects have several sensors simultaneously). The type, the placement, and the sampling rate of the sensor are varied through the each subject. This information is noted on the meta information file.

HASC2011corpus is composed from HASC2010corpus[3] and additional 20 subjects. Unlike the user supported activities like ALKAN[4], we carefully collect the data in the supervised matter to keep the quality of the activity data.

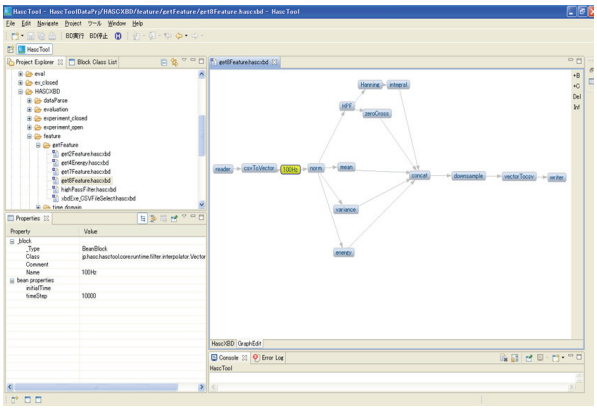

Figure 1. Large scale signal processing using HASC Tool

HASC TOOL

It is not easy to process a large number of data files for number of parameters. So we have built a tool for large scale automatic signal processing named "HASC Tool". By using HASC Tool, one can easily compose a task graph shown in Figure 1 to automate the evaluation process.

FINDINGS FROM HASC2011CORPUS

By using a large scale human activity corpus, we can perform various research tasks. As we already showed in [3], the effects of number of subjects are one of the most important results. In [3], we only perform the evaluation on 67 subjects. In HASC2011corpus, we employ additional 20 subjects. So we can show the further results on Table 2 and Figure 2. Details are on [3] and we add another 8th feature(integration of acceleration). As you can see in the Figure, recognition rates are still improving with the number of subjects. This result shows the further requirements of large scale corpus.

%	stay	walk	jog	skip	stUp	stDown
stay	86.7	4.4	2.0	0.7	2.2	3.9
walk	1.5	54.3	0.6	0.8	12.7	30.0
jog	1.3	0.9	49.9	32.4	0.2	15.3
skip	1.7	0.0	10.9	82.8	0.4	4.1
stUp	1.6	7.4	0.0	0.0	78.3	12.7
stDown	2.8	2.0	5.3	2.3	13.8	73.8
Overall	71.0					

Table 2. Confusion matrix of using 80 subjects from HASC2011corpus in the user-independent analysis.

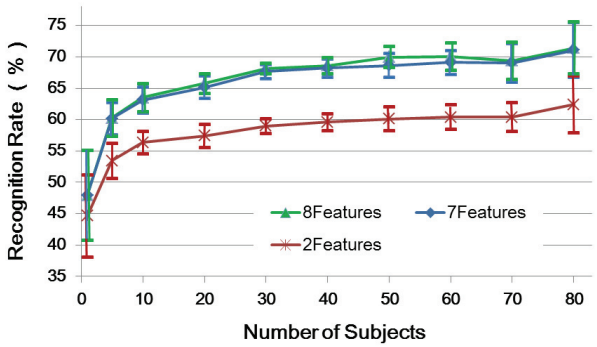

Figure 2. Effects of number of subjects in the user-independent analysis on the human activity recognition.

CONCLUSION AND FUTURE WORK

We have made a basic step for improving the human activity recognition field. Currently, we only have a dataset for simple activities. However, it is important to share the results on the common ground. Additionally, we have obtained a grant for gathering large scale activity dataset, so we can continue to gather human activity corpus to make and improve the common ground. We are also considering to gather more naturalistic activities.

ACKNOWLEDGMENTS

This work is partially supported by MEXT Grand-in-Aid for Scientific Research (A) 23240014 and JST CREST Kashin-Free project.

REFERENCES

1.Nobuo Kawaguchi, Nobuhiro Ogawa, Yohei Iwasaki, Katsuhiko Kaji, Tsutomu Terada, Kazuya Murao, Sozo Inoue, Yoshihiro Kawahara, Yasuyuki Sumi and Nobuhiko Nishio, HASC Challenge: Gathering Large Scale Human Activity Corpus for the Real-World Activity Understandings, *ACM Augmented Human (AH2011)*, pp. 27:1-27:5 (2011).

2.HASC Website: http://hasc.jp/en

3.Nobuhiro Ogawa, Katsuhiko Kaji, Nobuo Kawaguchi, Effects of Number of Subjects on Activity Recognition - Findings from HASC2010corpus -, *International Workshop on Frontiers in Activity Recognition using Pervasive Sensing(IWFAR2011),* pp. 48-51(2011).

4.Yuichi Hattori, Sozo Inoue, Go Hirakawa, A Large Scale Gathering System for Activity Data with Mobile Sensors, *IEEE International Symposium on Wearable Computers (ISWC2011)*,pp.97-100(2011).