Daily Activity Recognition Based on Acoustic Signals and Acceleration Signals Estimated with Gaussian Process

Masafumi Nishida[*], Norihide Kitaoka[†], and Kazuya Takeda[*]
[*]Nagoya University, Nagoya, Japan
E-mail: nishida@coi.nagoya-u.ac.jp Tel: +81-52-789-3626
[†]Tokushima University, Tokushima, Japan

Abstract— **We have created corpus of daily activities using wearable sensors. The corpus consists of sound and image data from a camera and motion signals from a smartphone for both indoor and outdoor activities over 72 continuous hours. We propose a method that can interpolate acceleration signals to any sample points with a Gaussian process in order to recognize daily activities. We conducted recognition experiments of daily activities using our corpus. Experimental results showed that the proposed method can improve recognition accuracy compared to a conventional method. This demonstrates the effectiveness of estimating acceleration signals with a Gaussian process to recognize daily activities.**

I. INTRODUCTION

There are many elderly people living alone, and computer and robotic technologies have great potential for assisting with the physical and mental effort required with certain activities in aging societies such as Japan. For example, if an elderly person has some sort of accident or emergency, it is possible to notice this early if an unusual action can be differentiated from a normal one by monitoring his/her action based on information technology. We aim to monitor the behavior of the elderly to support them in their everyday activities as well as in emergency situations. In both cases, it is important to recognize daily activities. Our focus here is acoustic and acceleration signals obtained from devices that are easy to wear in order to recognize daily activities.

There have been studies on audio signals in living environments. Stager et al. [1] recognized sounds from four groups – kitchen, office, workshop, and outdoors– based on a k-nearest neighbor classifier. Harma et al. [2] conducted an experiment with an acoustic surveillance system comprised of a computer and microphone situated in a typical office environment. Peng et al. [3] classified healthcare-related audio events using hidden Markov models and hierarchical hidden Markov models to monitor patients and the elderly. Mesaros et al. [4] presented a detailed evaluation of an HMM-based event detection and classification system using recordings of ten different natural environments. Espi et al. [5] presented an acoustic event detection and classification method that learns features from spectrogram patches with a deep neural network. Imoto et al. [6] proposed a method for estimating user activities by analyzing long-term acoustic signals represented as acoustic event temporal sequences.

Acceleration signals have also been used to recognize human activities. Iso and Yamazaki [7] extracted feature vectors by analyzing sensor data obtained from a cell phone and classified gaits with the Kohonen self-organizing map. Kwapisz et al. [8] evaluated a system that uses phone-based accelerometers to recognize daily activity using decision trees, logistic regression, and multilayer neural networks. Khan et al. [9] proposed a hierarchical-recognition scheme featuring state recognition at a lower level and activity recognition at an upper level. Zhu and Sheng [10] implemented a neural network for gesture spotting and a hierarchical hidden Markov model for context-based recognition. Khan et al. [11] presented the architecture and implementation of a smartphone position-independent activity recognition system. Cho et al. [12] and Zhu and Sheng [13] proposed a recognizer that works with an acceleration sensor and a GPS on a mobile device and estimates a user’s means of migration. Lane et al. [14] and Ouchi and Doi [15] reported on a system using acoustic and acceleration signals collected with a microphone and a mobile phone.

However, these studies focused on only a few daily activities, which were imitation actions and are not realistic activities. Furthermore, it is difficult to measure all sample points of acceleration signals accurately with motion. However, these studies have not addressed these problems. Therefore, we have created a corpus consisting of both indoor and outdoor daily activities using a small video camera and smartphone over 72 continuous hours [16]. In this paper, we developed a method that can interpolate acceleration signals to any segment with a Gaussian process. We also conducted recognition experiments of daily activities using acoustic and acceleration features and adopted a method using weighted likelihood based on a Gaussian mixture model (GMM) to recognize daily activities using multiple features.

This paper is organized as follows. Section 2 introduces our corpus of daily activities, and Section 3 describes our proposed estimation method of acceleration signals based on a Gaussian process and a recognition method of daily activities. In Section 4, we discuss the recognition experiments of daily activities. We conclude with a brief summary and mention of future work in Section 5.

II. Corpus of Daily Living Activity

The total duration of indoor and outdoor activities is shown in Fig. 1. "Washing" denotes washing clothes using a washing machine, "cleaning-bath" denotes cleaning the bathroom, "cleaning-room" denotes cleaning a room with a vacuum cleaner, "drying-clothes" denotes drying clothes inside the home, "brushing-teeth" denotes brushing teeth after a meal, "bicycle" denotes riding a bicycle, "cleaning-table" denotes clearing up and washing dishes after a meal, "bath" denotes taking a bath, "shopping" denotes shopping in convenience stores, supermarkets, and other shops, "car" denotes traveling in a car, "toilet" denotes using the toilet, "cooking" denotes cutting, boiling, and roasting food in a kitchen, "tv" denotes watching a TV program, "meal" denotes eating a meal at home or at a restaurant, "drive" denotes driving a car, "notepc" denotes using a notebook PC at home or at the office, "reading" denotes reading books, "office" denotes desk work etc., "smartphone" denotes reading and writing messages and browsing Web sites using a smartphone, "sleeping" was not included in the graph because its duration was very long (1,254 min.).

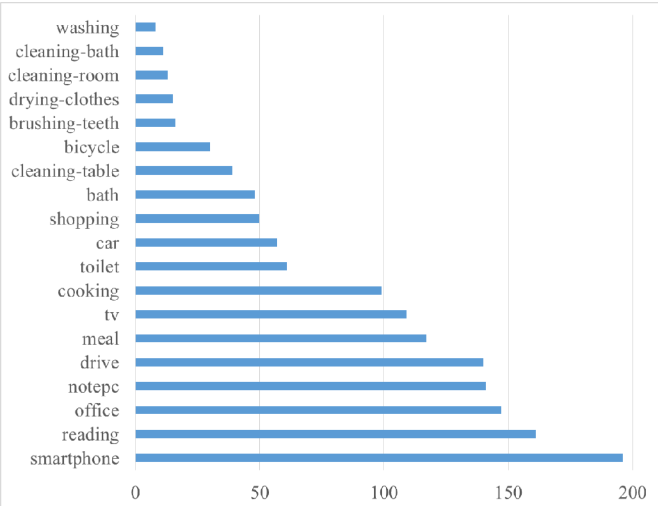

Fig. 1. Total duration of daily living activities (min)

III. Daly Living Activities Recognition

A. *Estimation of Acceleration Signals Based on Gaussian Process*

Gaussian processes can conveniently be used for Bayesian supervised learning, such as regression and classification [17]. We now describe Gaussian process methods for regression problems.

We have a dataset D of n observations, $D = \{(x_i, y_i) \mid i = 1, \ldots, n\}$, where x denotes an input vector of dimension D and y denotes a scalar output; the column vector inputs for all n cases are aggregated in the matrix X.

It is typical for more realistic modelling situations that we do not have access to function values, but only noisy versions in Equation (1), where K is called a covariance function or kernel. We used the kernel trick, which can be lifted into feature space by replacing occurrences of inner products.

$$y_i = f(x_i) + \varepsilon_i \qquad (1)$$
$$f \sim GP(\cdot \mid 0, K)$$
$$\varepsilon_i \sim N(\cdot \mid 0, \sigma^2)$$

We can write the joint distribution of the observed target values and the function values at the test locations under the prior as

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X,X)+\sigma^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right) \qquad (2)$$

The term marginal likelihood refers to the marginalization over the function values f. Under the Gaussian process model, the prior is Gaussian and a log marginal likelihood is

$$\log p(y \mid X, \theta) = -\frac{1}{2} y^T (K+\sigma^2 I)^{-1} y - \frac{1}{2}\log\left|K+\sigma^2 I\right| - \frac{n}{2}\log 2\pi. \qquad (3)$$

This result can be obtained directly by observing that $y \sim N(0, K+\sigma^2 I)$. In this study, we used a Gaussian kernel as the kernel function. The hyperparameters of the Gaussian kernel are estimated based on maximum likelihood theory.

$$k(x, x') = v^2 \exp\left(-\frac{(x-x')^2}{2r^2}\right) \qquad (4)$$

We applied the Gaussian process to the acceleration signals collected with the HASC Logger [18]. The acceleration signals were sampled at 200 Hz. However, data were observed in less than 200 samples in 1 second. Therefore, we interpolated the acceleration signals to 200 samples in 1 second with the Gaussian process. First, the hyperparameters of the kernel function were trained for every second using the acceleration signals collected with the HASC Logger. Second, the acceleration signals with 200 samples in 1 second were estimated based on the trained kernel function. These processes were applied for every second to each 1-minute segment in all daily living activities.

The average of samples of acceleration signal samples was 10,011 per minute in our corpus of daily activities. This shows that about 17% of sample points was not observed among the data sampled at 200 Hz. However, we were able to estimate the acceleration signals of 12,000 sample points to each 1-minute segment with the proposed method. Figures 2 and 3 show examples of acceleration signals with the HASC Logger and proposed method respectively, where x denotes a motion of right and left motion, y denotes movement toward the upper and lower sides, and z denotes motions in front and behind. Some sample points of acceleration signals were not measured in Figure 2. The proposed method could accurately estimate the acceleration signals in Figure 3.

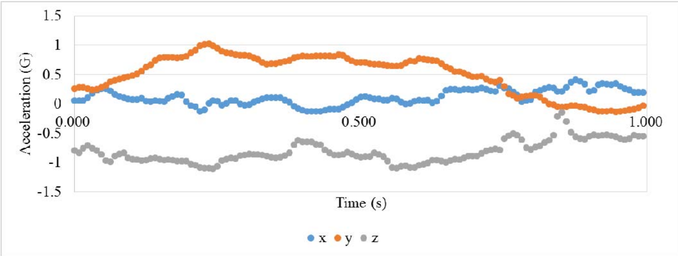

Fig. 2. Example of acceleration signals with HASC Logger (Bicycle)

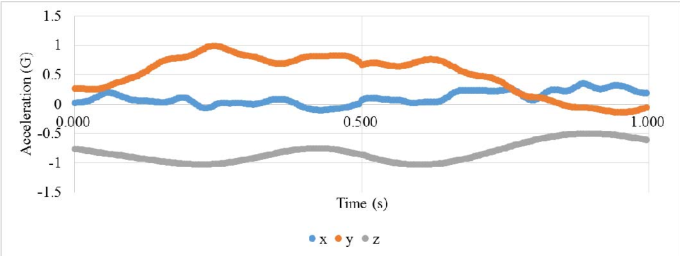

Fig. 3. Example of acceleration signals with proposed method (Bicycle)

We demonstrated that it is effective to interpolate the acceleration signals to any sample using a Gaussian process. However, the data of amount of 13 minutes were omitted from the following activity recognition experiments because the acceleration signals of these data were not estimated.

B. *Activity Recognition Based on Acoustic and Acceleration Signals*

We used two GMMs, one trained with acoustic features and the other with acceleration features, to recognize daily activities on the basis of the weighted log likelihood calculated by

$$\arg\max_{i} \frac{1}{N}\sum_{t=1}^{N}\left\{\log P(s_t \mid \lambda_s^i) + \log P(a_t \mid \lambda_a^i) \times w\right\}, \quad (5)$$

where N is the frame number of a segment, s is a t-th acoustic feature of a segment, a is a t-th acceleration feature of a segment, and w is a weight of likelihood.

IV. Activity Recognition Experiments

A. *Experiment setup*

We conducted recognition experiments on daily activities using acoustic signals and acceleration. Fourteen activities were used as candidates for recognition: “bicycle”, “cleaning-table”, “shopping”, “car”, “toilet”, “cooking”, “tv”, “meal”, “drive”, “notepc”, “reading”, “office”, “smartphone”, and “sleeping”. The acoustic features consisted of 24 Mel-frequency cepstral coefficients and log power. The total number of acoustic feature dimensions was 25. We compared two types of acceleration features. The first type was the averaged-value and a standard deviation for every second for each 1-minute segment in the data collected with the HASC Logger. The total number of acceleration feature dimensions was six. These features were mainly used to recognize activities. The second type was the acceleration signals estimated with the proposed method. Both acoustic and acceleration features were extracted every second without window overlap.

Both GMMs were trained using ten segments from each activity. The total duration of training data was ten minutes. The evaluation data included ten segments from each activity. The total number of evaluation data segments was 140. The evaluation data differed from the training data and were not included in the training data.

B. *Experimental results*

We conducted a recognition experiment with a varied mixture number of GMMs. The recognition accuracies for all combinations using only acoustic features are shown in Fig. 4. The maximum number of mixtures was seven because there were activities in which the weight of distribution could not be estimated when eight or more were used. The acoustic features obtained the highest accuracy of 72.3% when the number of mixtures was seven. In contrast, the acceleration features obtained the highest accuracy when there was only one mixture because there were activities in which the weight of distribution could not be estimated when two or more mixtures were used.

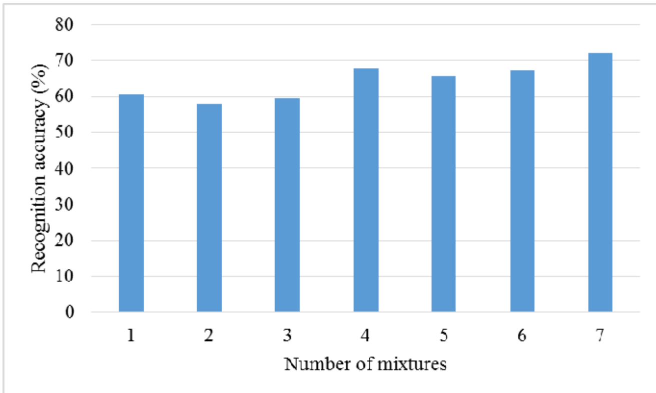

Fig. 4. Recognition accuracy using acoustic features only

Next, we conducted a recognition experiment using both acoustic and acceleration features. The recognition accuracies for every weight of likelihood are shown in Fig. 5. “Baseline” is a result from acceleration features in the data collected with the HASC Logger. “Gaussian process” is a result from acceleration features estimated with the proposed method. The number of mixtures of acoustic and acceleration features was seven and one, respectively. The recognition accuracy of the baseline was 75.7% when the weight was 0.3.

The recognition accuracy of the proposed method was 77.9% when the weight was 0.6. These results suggest that it is effective to use the acceleration features estimated with the proposed method in order to recognize daily activities. They also suggest that it is effective to add acceleration features into acoustic features based on the weight of likelihood.

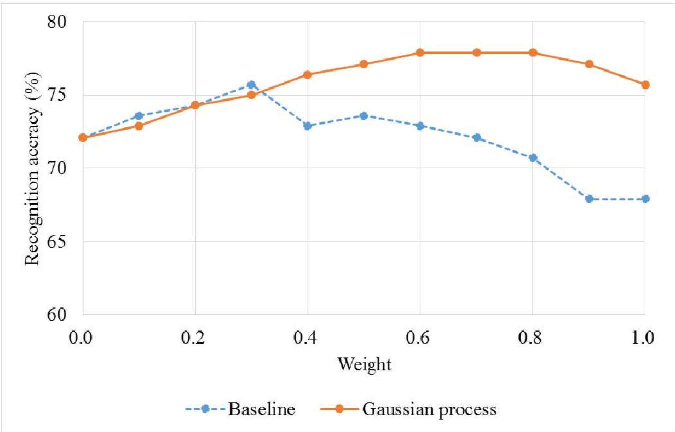

Fig. 5. Recognition accuracy for every weight of likelihood

V. CONCLUSIONS

We proposed a method that can interpolate acceleration signals to any sample points with a Gaussian process. Experimental results showed that the proposed method improved recognition accuracy compared with a conventional method. This demonstrates the effectiveness of estimating acceleration signals with a Gaussian process for recognizing daily activities.

For future work, we will conduct additional recognition experiments of daily activities using other machine learning methods and sensor signals. We also intend to create a larger corpus of daily activities consisting of data from several individuals.

ACKNOWLEDGMENT

I would like to express my gratitude to Mr. Yusuke Adachi for helpful experiments.

REFERENCES

[1] M. Stager, P. Lukowicz, N. Perera, T.Von Buren, G. Troster, and T. Starner, "SoundButton: Design of a Low Power Wearable Audio Classification System", Proceedings of the IEEE International Symposium on Wearable Computers (ISWC), pp. 12 -17, 2003.

[2] A. Harma, M. F. McKinney, and J. Skowronek, "Automatic Surveillance of the Acoustic Activity in Our Living Environment", Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), 2005.

[3] Y.-T. Peng, C.-Y. Lin, M.-T. Sun, and K.-C. Tsai, "Healthcare Audio Event Classification using Hidden Markov Models and Hierarchical Hidden Markov Models", Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), pp. 1218 -1221, 2009.

[4] A. Mesaros, T. Heittola, A. Eronen, and T. Virtanen, "Acoustic Event Detection in Real Life Recordings", Proceedings of the European Signal Processing Conference (EUSIPCO), pp. 1267 -1271, 2010.

[5] M. Espi, M. Fujimoto, D. Saito, N. Ono, and S. Sagayama, "A Tandem Connectionist Model Using Combination of Multi-scale Spectro-temporal Features for Acoustic Event Detection", Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4293 -4296, 2012.

[6] K. Imoto, S. Shimauchi, H. Uematsu, and H. Ohmuro, "User Activity Estimation Method Based on Probabilistic Generative Model of Acoustic Event Sequence with User Activity and Its Subordinate Categories", Proceedings of Annual Conference of the International Speech Communication Association (INTERSPEECH), pp. 2609 -2613, 2013.

[7] T. Iso and K. Yamazaki, "Gait Analyzer Based on a Cell Phone with a Single Three-axis Accelerometer", Proceedings of Mobile HCI, pp. 141 -144, 2006.

[8] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity Recognition using Cell Phone Accelerometers", ACM SIGKDD Explorations Newsletter, Volume 12, Issue 2, pp. 74 -82, 2010.

[9] A. M. Khan, Y.-K. Lee, S. Y. Lee, and T.-S. Kim, “A Triaxial Accelerometer-Based Physical-Activity Recognition via Augmented-Signal Features and a Hierarchical Recognizer”, IEEE Transactions on Information Technology in Biomedicine, Vol. 14, No. 5, pp. 1166 -1172, 2010.

[10] C. Zhu and W. Sheng, "Wearable Sensor-Based Hand Gesture and Daily Activity Recognition for Robot-Assisted Living", IEEE Transactions on Systems, Man, and Cybernetics, 2011.

[11] A. M. Khan, Y.-K. Lee, S. Y. Lee, and T. -S. Kim, "Human Activity Recognition via An Accelerometer-Enabled-Smartphone Using Kernel Discriminant Analysis", Proceedings of the IEEE International Conference on Future Information Technology (FutureTech), pp. 1 -6, 2010.

[12] K. Cho, N. Iketani, H. Setoguchi, and M. Hattori, "Human Activity Recognizer for Mobile Devices with Multiple Sensors", Proceedings of the Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing, pp. 114 -119, 2009.

[13] C. Zhu and W. Sheng, "Motion- and Location-based Online Human Daily Activity Recognition", Journal of Pervasive and Mobile Computing, 2011.

[14] N. D. Lane, M. Mohammod, M. Lin, X. Yang, H. Lu, S. A. Doryab, E. Berke, T. Choudhury, and A. T. Campbell, "BeWell: A Smartphone Application to Monitor, Model and Promote Wellbeing", Proceedings of the International ICST Conference on Pervasive Computing Technologies for Healthcare, 2011.

[15] K. Ouchi and M. Doi, “Living Activity Recognition using Off-the-shelf Sensors on Mobile Phones“, Annals of Telecommunications, Vol. 67, pp. 387 -395, 2012.

[16] M. Nishida, N. Kitaoka, and K. Takeda, “Development and Preliminary Analysis of Sensor Signal Database of Continuous Daily Living Activity over the Long Term”, Proceedings of APSIPA, pp. 1 –6, 2014.

[17] C. E. Rasmussen and C. Williams, “Gaussian Processes for Machine Learning”, MIT Press, 2006.

[18] N. Kawaguchi, H. Watanabe, T. Yang, N. Ogawa, Y. Iwasaki, K. Kaji, T. Terada, K. Murao, H. Hada, S. Inoue, Y. Sumi, Y. Kawahara, and N. Nishio, “HASC 2012 corpus: Large Scale Human Activity Corpus and Its Application”, Proceedings of the Second International Workshop of Mobile Sensing: From Smartphones and Wearables to Big Data, pp. 10 -14, 2012.