Proceedings of 2009 APSIPA Annual Summit and Conference, Sapporo, Japan, October 4-7, 2009

Human Action Recognition Using Acceleration Information Based On Hidden Markov Model

Shin'ichi Takeuchi* Satoshi Tamura† and Satoru Hayamizu†
* Virtual System Laboratory, Gifu University, 1-1 Yanagido, Gifu 501-1193, Japan
† Faculty of Engineering, Gifu University, 1-1 Yanagido, Gifu 501-1193, Japan
E-mail: takeuchi@hym.info.gifu-u.ac.jp

***Abstract*—This paper investigates the best feature parameter for human action recognition by using Hidden Markov Model (HMM) with triaxial acceleration sensor information. Our target is to recognize six types of basic actions (walk, stay, sit down, stand up, lie down, get up) in the room of daily life. First, as parameters in time domain, acceleration information in three axes and their derivatives are used as the baseline method. Secondly, Mel-Frequency Cepstral Coefficients are used as feature parameters in frequency domain. As the recognition result, the best recognition rate was obtained when MFCC and its Δ elements of the axis that contained most of gravitational acceleration information.**

I. Introduction

Recently, many works about recognition of human activity in daily living (ADL: Activity in Daily Living) have been done [1]–[10]. Henry [1] proposed automatic detection, tracking, and recognition system in office environment using video data. Also, there are many researches about human action recognition. Small acceleration sensor is used widely as one of the action measurement methods using wearable sensor. Sometimes it is used with other type of sensors. Triaxial acceleration sensor is often used to recognize ADL. Jin [2] used a wireless wearable sensor network for monitoring ADL. Krishnan [3] compared the recognition performance of AdaBoost, Support Vector Machine (SVM), and regularized logistic regression. Song [4] and Karatonis [5] used a waist-mounted sensor for recognition. Lee [6] used acceleration and angular velocity data for recognition and classification.

Various recognition techniques use the information from the sensor. Typical pattern classification such as SVM [7] or Neural Network [8] was investigated. Hidden Markov Model (HMM) is one of the effective pattern recognition methods [2], [9], [10]. Wu [9] proposed a real-time classification system which uses HMMs and an extended Kalman filter to acceleration information. By Ward [10], sound and acceleration information were used to recognize ADL, applying linear discriminant analysis and HMMs respectively.

We have developed support system of human life in the room [11]–[13]. As effective machine learning and pattern recognition algorithm, we have used HMM for human action recognition in daily living. However, it is necessary to find out the best feature parameters and model conditions.

The purpose of this paper is to clarify the optimal feature parameters of acceleration sensor for human action recognition. As feature parameters in time domain, acceleration data and their time derivatives (called delta-parameters in speech recognition) are used. As feature parameters in frequency domain, Mel-Frequency Cepstral Coefficients (MFCC) are used.

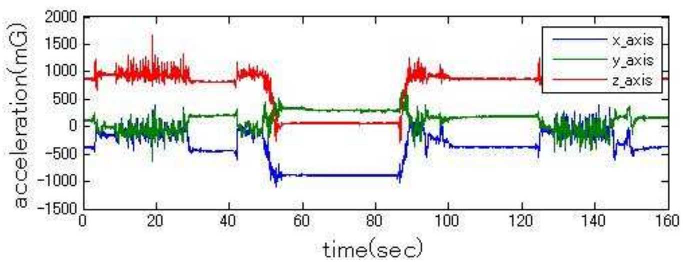

Fig. 1. Acceleration data.

Experiments are conducted in order to verify the effectiveness of these parameters and to find out optimal conditions for feature types, dimension of parameters, HMM model size, and combination of axes.

The rest of this paper is organized as follows: Section II introduces the triaxial acceleration sensor and our action recognition method. In Section III, several acceleration features are proposed. Experiments are conducted and discussion is mentioned in Section IV and V respectively. Finally Section VI concludes this paper.

II. Recognition method

The acceleration sensor device used in this paper is a small-sized wireless acceleration sensor made by Wireless Technology Inc. It measures acceleration information on three axes (X, Y, and Z axis). The acceleration sensor is put on human's waist (in front of a belt), and measures each 10ms (the sampling frequency is 100Hz). An example of actual acceleration signal is shown in Fig. 1. The vertical axis shows acceleration amplitude and the horizontal axis shows the time. The gravitational acceleration is included in the measured data. The acceleration data is transmitted to the host computer using Bluetooth. The maximum range is $\pm$ 3G and resolution is 8.8mG.

HMM is a statistical modeling method that is often used in the speech and image recognition. Also, HMM is used to develop the finger motion recognition with a three-dimensional

Fig. 2. Environment of data measurement.

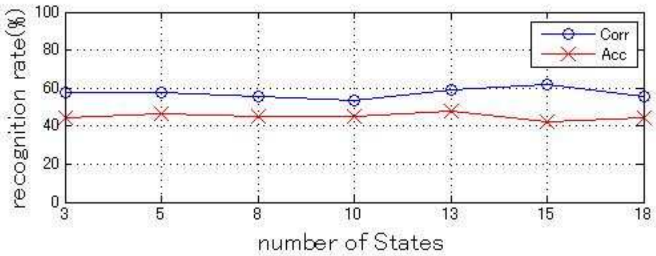

Fig. 3. Result of 6 dimensions.

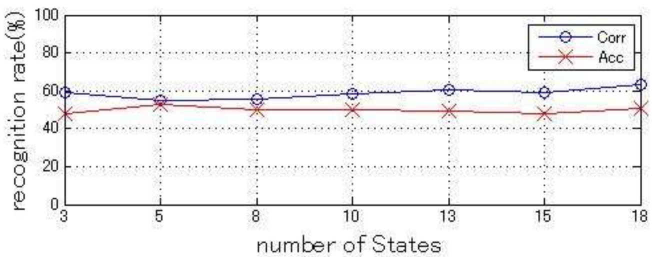

Fig. 4. Result of 9 dimensions.

positional sensor [14]. Therefore, HMM is expected to recognize the action effectively.

In this paper, continuous and left-to-right type HMM is used. For experiments, HMM Tool Kit (HTK) is used.

III. Action data

The target action patterns (six types in total) are shown as follows.

- walk
- stay
- sit down
- stand up
- lie down
- get up

These patterns are considered as the basic actions in daily life. The experimental data is obtained from three male adults, age of 20-30. The environment of data measurement is shown in Fig. 2. The testees put an acceleration sensor on front of waist.

Training and evaluation are done by different data sets. Four data sets from one testee are used as the training data (each action is from 5 to 7 repetitions in data). Data sets obtained from two testees are used as the evaluation data (open condition). Under closed evaluation condition, the best recognition results were 95.75% in correct and 93.74% in accuracy. The feature parameter used for the action recognition is these acceleration data, and is examined the best condition of the action recognition.

IV. Parameters in time domain

First of all, feature parameters in time domain are examined [13]. As the baseline method, two types of feature parameters are used. The one of them is triaxial acceleration data and its Δ elements (total 6 dimensions), and another is triaxial acceleration, Δ, and its $\Delta\Delta$ elements (total 9 dimensions). The delta (Δ) parameters are time derivations of acceleration data and widely used in speech recognition. Each Δ elements are calculated as linear regression coefficient with window size of 100 points (1 sec).

Also, the number of states of HMM is changed across 5, 8, 10, 13, 15, and 18, and the influence on recognition is examined. Fig. 3 and Fig. 4 show each result. Here, the performance of the action recognition is evaluated by Correct (Corr) and Accuracy (Acc).

$$\text{Corr} = \frac{H}{N}$$

$$\text{Acc} = \frac{H - I}{N}$$

N is total number of action, H is amount of correct, and I is insertion error. These results are around 60% in both dimension, and have small differences. It is considered that $\Delta\Delta$ elements does not contribute to action recognition in time domain. Also, when the number of states is changed, recognition rate does not change dynamically.

V. Parameters in frequency domain

To improve the recognition performance from the result in Section IV, other feature parameter is examined. Each action is analyzed with the frequency to see characteristics of acceleration information.

Fig. 5 shows an acceleration example for “lie down” and Fig. 6 shows an example of frequency analysis for the data. Most of energy exists in the low frequency, and it resembles the feature of the voice. For those reasons, the improvement of recognition performance can be expected by using MFCC, which is mostly used in speech recognition as feature parameters.

In this paper, the best condition is examined by changing the combination of axes and the number of dimension of MFCC as the feature parameters. The conditions of calculating MFCC is as follows: frame length is 100 points (1 sec), and frame shift is 50 points (0.5 sec).

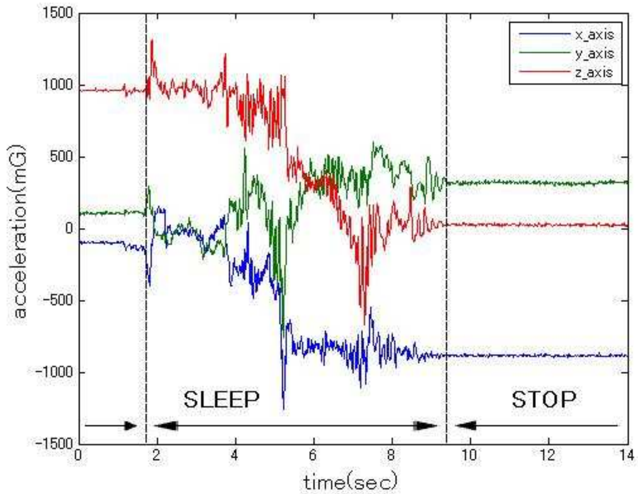

Fig. 5. Acceleration data example (lie down).

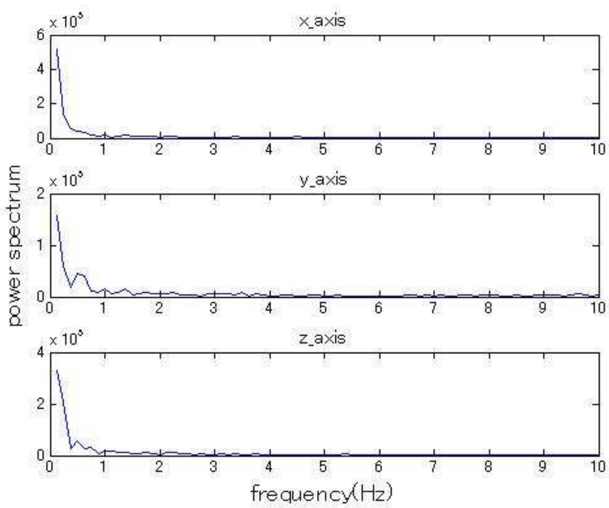

Fig. 6. Frequency analysis example (lie down).

A. Combination of axes

X, Y, and Z axis of the acceleration information are clearly different, and it is not necessarily for the best condition to use all data of the three axes like Section IV. Then, the best combination of the three axes is examined using MFCC as the feature parameter. The following seven combinations are examined.

- X axis only
- Y axis only
- Z axis only
- X and Y axis
- Y and Z axis
- X and Z axis
- X, Y, and Z axis

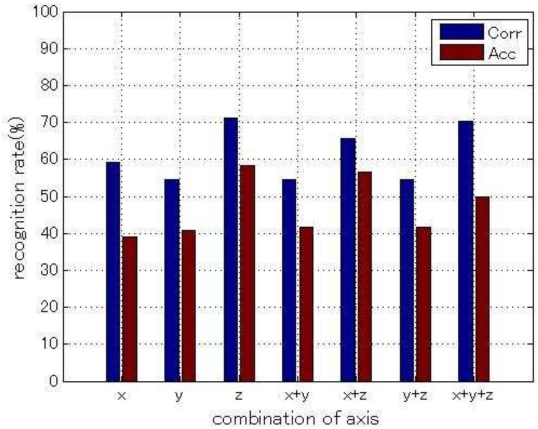

Fig. 7. The case of changing the combination of axes.

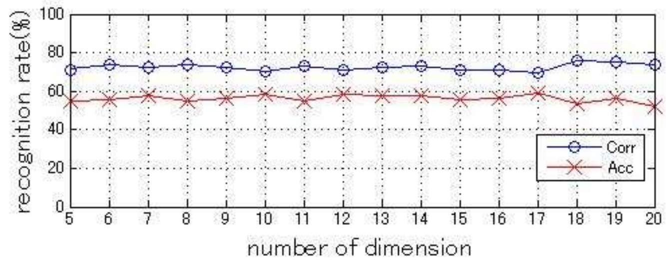

Fig. 8. The number of dimension and recognition rate.

Fig. 7 shows the result of the above combinations. This result shows that only with Z axis is the best. For action recognition, it should use acceleration information including most of gravitational acceleration information. Therefore, only Z axis is used in following experiments.

B. Number of dimension

Following experiments are for the number of dimension of MFCC. The experiments in previous section used 12 dimensions. In order to optimize the condition of frequency analysis, the number of dimension is changed from 5 to 20 dimensions and the best number of dimension is examined. Fig. 8 shows the result of each dimension.

The correct rate and accuracy were almost same overall from 5 to 15 dimension. In this feature parameter condition, it can be said that the change by the dimension hardly affects.

C. Adding frame level derivatives

Next, improvement of recognition performance is attempted by adding derivative parameters in frame level. First of all, feature parameter of Δ elements are added. New feature parameter created by adding each Δ elements to the MFCC (6, 12 and 18 dimensions). Each Δ elements are calculated using 10 frames.

Fig. 9 shows the comparison with and without Δ elements for each dimensions case. The result from 6 dimensions improves to 80% by using Δ elements. Also entire results

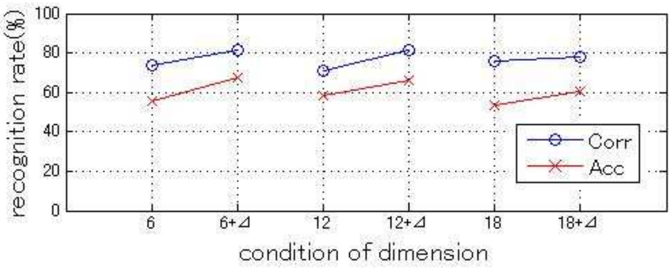

Fig. 9. Effectiveness of adding the time derivatives.

show that using Δ elements brings improvement of recognition rate.

D. Discussion

Summarizing these results, using MFCC (frequency domain) was better than the acceleration data (time domain). Also adding the time derivatives (delta) produces improvements.

For combination of axes, acceleration information only in one vertical axis was the best. It is thought that gravitational information possesses most significance for discrimination of human action in our case.

The number of dimension of MFCC affected the recognition performance slightly. As shown in the frequency analysis of Fig. 6, energy in low frequency is dominant. Feature parameters other than MFCC should be investigated.

Also, MFCC is calculated using biased filter bank, but it should be necessary to use more suitable filter bank for action recognition. Finally, acceleration sensor data contain both action and posture information. This is indicated by the result that only one axis with most of gravitational acceleration information.

VI. Conclusion

This paper presents action recognition using HMM with acceleration sensor information as feature parameters. In frequency domain, MFCC brought better recognition result than time domain parameters. And frame level time derivatives (delta) also produced improvement in recognition performance. For combination of axes, an axis with most of gravitational acceleration information was the most important for action recognition. As future work, robustness for the change of human subjects and action types should be studied. We expect speech recognition techniques will be effective for these purposes.

References

[1] T. C. C. Henry, E. G. Ruwan Janapriya, and L. C. de Silva, "An automatic system for multiple human tracking and actions recognition in office environment,", *Proc. of ICASSP2003,* vol. 3, pp. 45–8, Apr. 2003.

[2] He Jin, Li Huaming, and Tan Jindong, "Real-time Daily Activity Classification with Wireless Sensor Networks using Hidden Markov Model", *Proc. of EMBS,* pp. 3192–3195, Aug. 2007.

[3] N. C. Krishnan and S. Panchanathan, "Analysis of low resolution accelerometer data for continuous human activity recognition," *Proc. of ICASSP2008,* pp. 3337–3340, Mar. 2008.

[4] Sa-kwang Song, Jaewon Jang, and Soojun Park, "A Phone for Human Activity Recognition Using Triaxial Acceleration Sensor," *Proc. of ICCE2008,* pp. 1–2, Jan. 2008.

[5] D. M. Karatonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler, "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring," *IEEE Trans. on Info. Tech. in Biomedicine,* vol. 10, no. 1, pp. 156–167, 2006.

[6] Seon-Woo Lee and K. Mase, "Activity and location recognition using wearable sensors,", *IEEE Pervasive Computing,* vol.1, no.3, pp. 24–32, 2002.

[7] Zhen-Yu He and Lian-Wen Jin, "Activity recognition from acceleration data using AR model representation and SVM," *Proc. of MLC2008,* vol. 4, pp. 2245-2250, July 2008.

[8] M. Ermes, J. Parkka, J. Mantyjarvi, and I. Korhonen, "Detection of Daily Action and Sports With Wearable Sensors in Controlled and Uncontrolled Conditions,", *IEEE Trans. on Inf. Tech. in Biomedicine,* vol. 12, no. 1, pp. 20–26, 2008.

[9] Jian Kang Wu, Liang Dong, Wendong Xiao, "Real-time Physical Activity classification and tracking using wearable sensors,", *Proc. of ICIC2007,* pp. 1–6, 2007.

[10] J. A. Ward, P. Lukowicz, G. Troster, and T. E. Starner, "Activity Recognition of Assembly Tasks Using Body-Worn Microphones and Accelerometers," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 28, no. 10, pp. 1553–1567, 2006.

[11] H. Niwa, T. Kawamura, S. Tamura, and S. Hayamizu, "Automatic indexing and alignment of videos for medical care training," *Proc. of VSMM2007,* pp. 245–253, Sep. 2007.

[12] S. Itou, S. Asano, S. Tamura, and S. Hayamizu, "Human motion detection with 3D-accelerometers using statistical voice activity detection method," *Proc. of IPSJ2008,* vol.4, pp.225–226, 2008 (in Japanese).

[13] S. Asano, S. Itou, S. Takeuchi, S. Tamura, and S. Hayamizu, "Behavioral Distinction Based On Hidden Markov Model Using Triaxial Acceleration Sensor," *Proc. of FIT2008,* 2008 (in Japanese).

[14] K. Morimoto, C. Miyajima, N. Kitagawa, K. Itou, and K. Takeda, "Statistical Segmentation and Normalization of Trajectories of Finger-Tip Movements for a Gesture Interface," *IEICE,* SP2007-45, 2007 (in Japanese).