HAAR-LIKE FILTERING FOR HUMAN ACTIVITY RECOGNITION USING 3D ACCELEROMETER

Yuya Hanai, Jun Nishimura and Tadahiro Kuroda

Department of Electronics and Electrical Engineering, Keio University,
3-14-1, Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan
{hanai, nishimura}@kuro.elec.keio.ac.jp, kuroda@elec.keio.ac.jp

ABSTRACT

In this paper, novel 2 one-dimensional (1D) Haar-like filtering techniques are proposed as a new and low calculation cost feature extraction method suitable for 3D acceleration signals based human activity recognition. Proposed filtering method is a simple difference filter with variable filter parameters. Our method holds a strong adaptability to various classification problems which no previously studied features (mean, standard deviation, etc.) possessed. In our experiment on human activity recognition, the proposed method achieved both the highest recognition accuracy of 93.91% while reducing calculation cost to 21.22% compared to previous method.

Index Terms**—** 1D Haar-like filtering, human activity recognition, accelerometer, sensornet

1. INTRODUCTION

Acceleration signals obtained from wearable sensors contain valuable information originated from human activity. Considering the practical sensornet applications such as our work of "Business Microscope" [1] (Figure 0), in which sensor node is driven by strictly limited power and run on the low duty ratio in the order of 0.1, only classified results processed in the sensor node can be sent to server once a few second. This is because transmitting raw signals consumes a lot of energy. Therefore, low calculation cost yet highly reliable human activity recognition system is needed.

However, most of previous works on human activity recognition using accelerometer did not focus on calculation cost. In these works, same features such as mean, standard deviation, energy, correlation and frequency-domain entropy are adopted [2], [3], [4]. Some of them require FFT. In [5], they concluded FFT coefficients are effective for the classification. Except for mean, large amount of multiplication is needed to calculate these feature values. Furthermore, since these methods are not specifically designed to extract features from acceleration signals which can contribute to human activity recognition, there are many redundancies in features themselves. This can cause inefficient classifier to be generated by training with such features.

To alleviate these problems, this paper presents 1D Haar-like filtering and 1D Haar-like biaxial filtering as a totally new feature extraction method for 3D acceleration signals. For its flexibility, this simple filter can design itself to be able to extract features effective for various classifications. In addition, our method does not require any multiplication. This approach is originally inspired by the 2D Haar-like feature, which is used in state-of-the-art face detector for its high performance and calculation efficiency [6].

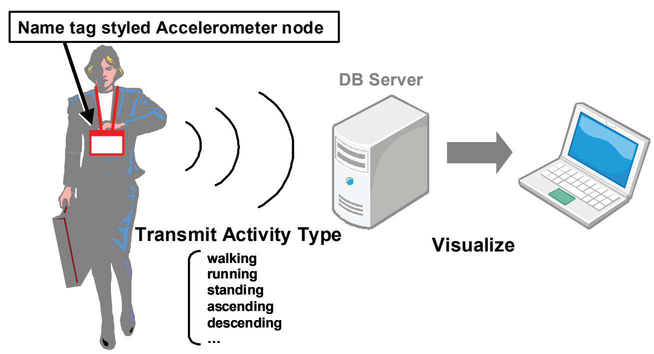

Figure 1. Supposed application.

2. PREVIOUS EXTRACTION METHODS

2.1. Activity Recognition

This section describes 4 major feature extraction methods for human activity recognition and discusses their calculation cost.

The feature extraction of "mean" is the lowest calculation cost among previous methods, whose actual computation equation is given by

$$mean = \sum_{t=1}^{W_{Frame}} s(t) \qquad (1)$$

where $s(t)$ denotes raw signal input value and W_{Frame} denotes filtering frame width, which is 512(2.56sec) in our experiment, thus 512 times addition is needed for the calculation per frame. In this application, division is not required because W_{Frame} is fixed value. Although the calculation cost is low, the modeling ability of activity is also low. This feature extracts "rough moving direction," but that's it.

The feature "standard deviation" has relatively low calculation cost of the four. The actual calculation of the feature "standard deviation" is given by

$$\sigma = \sqrt{\frac{1}{W_{Frame}} \sum_{t=1}^{W_{Frame}} \left(s(t) - s_{mean}\right)^2} \quad (2)$$

where

$$s_{mean} = \frac{1}{W_{Frame}} \sum_{t=1}^{W_{Frame}} s(t) \quad (3)$$

In this case, the actual computation is equivalent to its definition because removing square-root causes different distribution of the feature, thus generating different classifier. The calculation cost roughly corresponds to W_{Frame} times addition and W_{Frame} times multiplication because the division and the square-root operation are not dominant. This feature extracts "rough energy."

The feature "correlation" is calculated between each pair of axes as the ratio of the covariance and the product of the standard deviation, thus relatively high calculation cost.

$$corr(x,y) = \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y} \quad (4)$$

This feature is useful for classifying activities which involves more than one dimensional acceleration signal such as ascending and descending.

The feature "energy" is extremely high calculation cost method. It is the sum of the squared discrete FFT component magnitudes of the signal and given by

$$Energy = \frac{1}{W_{Frame}} \sum_{i=1}^{W_{Frame}} X_i^2 \quad (5)$$

where Xi denotes each FFT component. Even FFT, which is the order of O(W_{Frame} log W_{Frame}), is adopted, large amount of floating point multiplication such as exp(x) is required.

2.2. Image Recognition

In image recognition field, 2D Haar-like feature (Figure 2) is widely adopted. The feature value is defined by the difference of the sum of positive area and the sum of negative area in grayscale image.

$$featurevalue = \sum_{PositiveArea} image(x,y) - \sum_{NegativeArea} image(x,y)$$

(6)

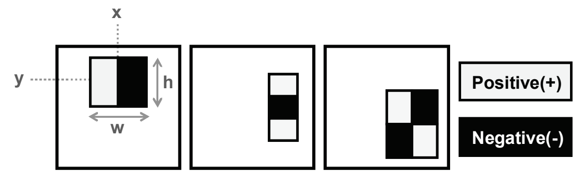

Figure 2. Some 2D Haar-like feature.

Naturally, it requires no multiplication. However, by using raw input data, the feature that has large area requires large amount of memory access and addition/subtraction that is equivalent to the area size. "Integral image" [6] enables the calculation using only up to 9 memory access per feature.

3. 1D HAAR-LIKE FILTERING

3.1. 1D Haar-like Filtering

To apply Haar-like filtering to acceleration signals, we used simple one dimensional filter as a basis with its filter width and shift width variable within a certain limitation. By controlling their degrees of freedom, various feature search spaces can be designed to fit to a certain recognition problem.

As shown in Figure 3, the feature is the sum of the absolute output of Haar-like filtered signals. This is calculated by the equation which is given by

$$x_m = \sum_{n=0}^{N} \left| \sum_{k=0}^{W_{Filter}} h_m(k) s(n \cdot W_{Shift} + k) \right| \quad (7)$$

where $s(t)$ denotes raw signal input value, $h_m(k)$ denotes m-th Haar-like filter function of M filters pool, W_{Filter} denotes filter width of the m-th Haar-like filter. N is given by

$$N = (W_{Frame} - W_{Filter}) / W_{Shift} + 1 \quad (8)$$

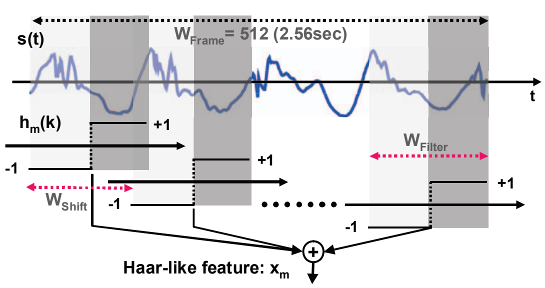

Figure 3. 1D Haar-like filtering for acceleration signals.

where W_{Shift} is shifting value, which takes 10 to 100% of W_{Filter}. Unlike the use of Haar-like filters in face detection shown in Figure 1, the position of the Haar-like filter relative to the analysis frame does not yield meaningful information. For 3D acceleration signals, this filtering is carried out for each axis respectively.

3.2. 1D Haar-like Biaxial Filtering

The use of 1D Haar-like filtering for each axis cannot measure the correlation between axes. According to [2], features that extract correlation between axes can improve the activity recognition result since the accelerometer data is tri-axial.

Therefore, we introduce "1D Haar-like biaxial filtering (Figure 4)" to measure the correlation between 2 axes for further improvement. The procedure is almost same as basic 1D Haar-like filtering. The feature is the sum of the absolute difference between 2 axial Haar-like filtered signals. This is calculated by the equation given by

$$xy_m = \sum_{n=0}^{N} \left| \sum_{k=0}^{W_{Filter}} h_m(k) X - \sum_{k=0}^{W_{Filter}} h_m(k) Y \right| \tag{9}$$

where X and Y denote raw signal input value of x and y axis respectively given by the following equation.

$$X = X\left(n \cdot W_{Shift} + k\right) \tag{10}$$

$$Y = Y\left(n \cdot W_{Shift} + k\right) \tag{11}$$

For 3D acceleration signal, this filter is carried out for each two axes respectively (*e.g.* xy, yz and zx. (9) to (11) are the case of xy). Using this filter, recognition problems on 3D signals can be solved much more efficiently.

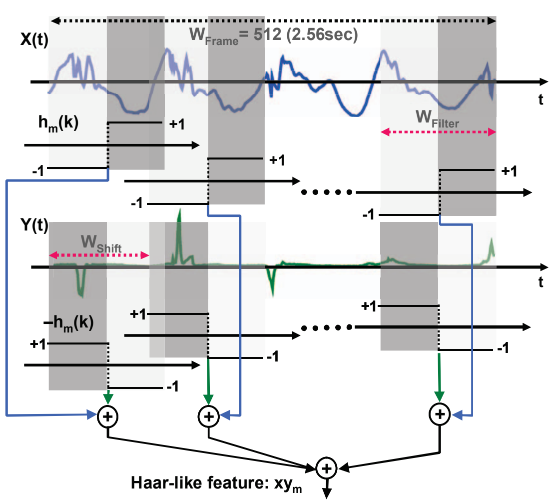

Figure 4. 1D Haar-like biaxial filtering for 3D signals.

3.3. Integral Signal

To further decrease the calculation cost of the Haar-like filtering, "integral signal [7]," which is integral image [6] application to one dimensional signal, is introduced.

Before calculation of the Haar-like filter, integral signal is generated as preprocessed intermediate signal, which is given by

$$is(t) = \sum_{t'=0}^{t} s(t') \tag{12}$$

where $s(t)$ is original input signal.

By using this integral signal, each Haar-like filtered value is calculated via only 3 point memory access. In other words, it requires only 1 addition, 1 subtraction and 1 bit-shift, while using original input signal requires W_{Filter} times addition per filter.

4. EXPERIMENT

4.1. Data Collection

The 200Hz 3D acceleration data was collected from 4 people for over 120 minutes using off-the-shelf wireless accelerometer marketed by ATR-Promotions. As shown in Figure 5, the accelerometer is bonded on name tag.

The data is transmitted to PC via Bluetooth and includes 5 actions:

- Walking
- Running
- Standing
- Ascending
- Descending

No post-processing was carried out and classes were labeled manually.

4.2. Feature Extraction and Generate Classifier

Various feature extractions were executed for comparison. We did the combinations of "M (mean)," "MS (mean+standard deviation (sd))," and "MSEC (mean+sd+energy+correlation)." As for Haar-like filtering, W_{Frame} was fixed to 512 (2.56sec), W_{Filter} was controlled from 1 to 512 and W_{Shift} was controlled from 10 to 100% of its filter width.

To generate classifier of decision tree with pruning (C4.5), the Weka Machine Learning Toolkit was used. 10-fold cross-validation was carried out for the evaluation.

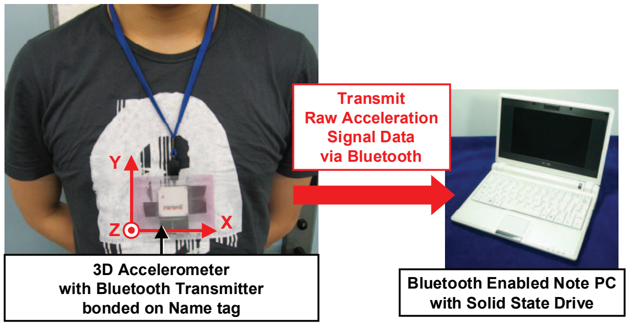

Figure 5. Data collection environment.

5. EXPERIMENTAL RESULTS

The relations between recognition accuracy and the calculation cost (total amount of add+multiplication operations according to the path of the decision tree taken by each test input, where 16bit short multiplication is estimated as 16*add operations by Booth's multiplication algorithm) are plotted in Figure 5. Among previous methods, "MS" achieves both high accuracy of 92.64% and low calculation cost of 16.61k operations per window (opw).

Using basic Haar-like filter (H), high accuracy of 92.69% is achieved with 3.08k opw, which is lowest among the methods with over 90% accuracy. By adding Haar-like biaxial filter (HHB), the highest accuracy of 93.91% is yielded with 3.53k opw.

The trained decision tree of "HHB" consists of only 65 nodes, while "H" contains 81 nodes and "MS" contains 83 nodes.

The confusion matrix of "HHB" is shown in Table 1.

5. CONCLUSIONS AND DISCUSSIONS

1D Haar-like filtering and 1D Haar-like biaxial filtering are proposed as a totally new feature extraction method for 3D acceleration signals based human activity recognition. The use of these 1D Haar-like filters achieved the highest accuracy of 93.91% with calculation cost of 21.22% compared to the previous method. With the help of 1D Haar-like biaxial filter, this method provided 17.8% less redundant trained classifier, since these filters have a strong adaptability gained by the high degrees of freedom.

Naturally, other learning algorithms such as Boosting and SVMs can be easily applied to Haar-like filter based framework, which is popular way in image recognition problem. Combined application of Haar-like filter and popular learning algorithm will enable recognition front-end unification (image, video, sound and acceleration signals).

6. REFERENCES

[1] J. Nishimura, *et al.*, "Speech 'Siglet' Detection For Business Microscope," *In Proceedings of PerCom*, 2008.

[2] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," *In Proceedings of PerCom,* 2004.

[3] P. Nurmi, *et al.*, "A framework for distributed activity recognition in ubiquitous systems," *In Proceedings of ICAI*, 2005.

[4] N. Ravi, *et al.*, "Activity Recognition from Accelerometer Data," *In Proceedings of AAAI*, 2005.

[5] T. Huynh and B. Schiele, "Analyzing features for activity recognition," *Proceedings of joint conference on Smart objects and ambient intelligence,* 2005.

[6] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," *In Proceedings of CVPR*, 2001.

[7] J. Nishimura, et al., "Speech Detection using Haar-like Filtering for Sensornet," *In Proceedings of ICST*, 2008.

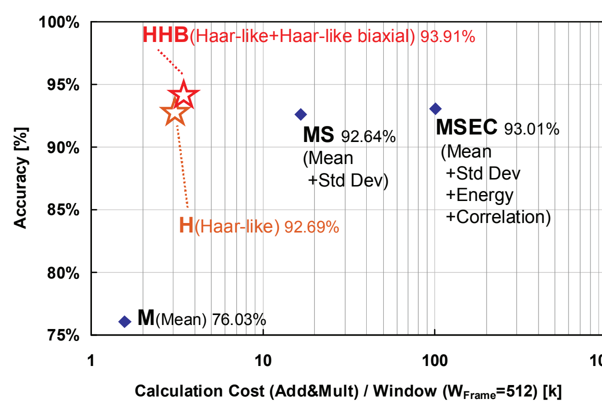

Figure 6. Calculation cost vs. recognition accuracy.

Table 1. Confusion matrix.

classified as →	walking	running	standing	ascending	descending
walking	848	0	7	39	14
running	0	909	0	0	0
standing	6	0	814	1	1
ascending	26	0	0	503	59
descending	18	1	0	58	471