ROTATION INVARIANT FEATURE EXTRACTION FROM 3-D ACCELERATION SIGNALS

Takumi Kobayashi, Koiti Hasida and Nobuyuki Otsu

National Institute of Advanced Industrial Science and Technology
Tsukuba, 1-1-1 Umezono, Japan

ABSTRACT

In this paper, we propose a method to extract features from three-dimensional acceleration signals. The proposed method is based on the (auto-)correlation matrix of Fourier transform features, naturally containing the correlations between the frequencies as well as the ordinary power spectrum for each frequency. The proposed features are inherently invariant to both rotational variations and temporal shift (delay), whereas the other methods employ ad hoc preprocessing to increase robustness to those variations. Thereby, we can favorably apply the proposed method to analyze 3-D acceleration signals regardless of the orientations of the accelerometer. In the experiment on gait identification using an accelerometer embedded in a cellular phone, the proposed method outperformed the other methods.

Index Terms— Three-dimensional acceleration signal, feature extraction, auto-correlation matrix, rotation invariance, gait recognition

1. INTRODUCTION

Three-dimensional accelerometers are commonly used sensors, as they have been widely embedded in cellular phones. Based on the signals of such an accelerometer, the orientation of the phone can be recognized and the screen is accordingly rotated so as to facilitate users' browsing. People carry the cell phones with such sophisticated functions in daily life. Keen attentions are therefore paid to the possibilities of using accelerometers for providing personalized services to improve the quality of life. For example, action recognition by accelerometers contributes to not only monitoring health but also selecting services according to the person's situations. The accelerometer would be also useful for identifying persons, especially based on human gaits. The identification plays a key role for security in any services. The gait identification by the accelerometer is recently studied [1, 2], while it has been mainly based on vision so far [3, 4].

The accelerometer measures acceleration vectors of the device along the three (x, y, and z) axes which are defined on the device. Along with the rotations of the device, the obtained signals are rotated across those axes (see Fig. 1). For dealing with the acceleration signals, it is hence important
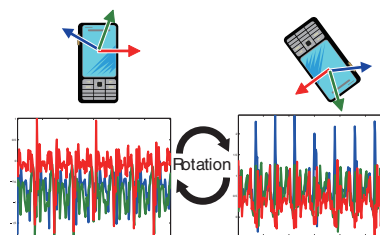

Fig. 1. Three-dimensional acceleration signals. The three (x, y, and z) axes are defined on the device, drawn in three colors (red, green, and blues). The signals are rotated along with the rotation of the device. (This figure is best viewd in color.)

to extract the essential features of human actions independently of the device orientations: that is, the extracted features should be invariant or at least robust to the rotations of the device.

For this purpose, some studies have been made so far. Most methods apply ad hoc approaches to achieve the rotation robustness. In [5], three-dimensional signal vectors are first normalized by applying principal component analysis (PCA), and then wavelet features are extracted from the normalized signals in each axis. PCA statistically normalizes the orientations of the signals. In [6], the gravity vector is statistically estimated as the mean vector of the three-dimensional signals. Then, simple statistics based on the gravity vector are extracted as features; such as the mean and the standard deviation of the inner product values between the input signal and the estimated gravity vectors. These statistical preprocessing methods to compensate the signal orientations require plenty of signal data for obtaining statistically stable results, and the rotation robustness of the features relies on the stability (accuracy) of the results. On the other hand, the method in [7] increases robustness in the process to learn the classifier by augmenting training samples with the simulated signals that are artificially rotated. To accomplish high robustness, significant augmentation is required and thereby the sample distribution becomes complex, making it difficult to train the classifier.

In this paper, we propose a method to extract *rotation invariant* features from the acceleration signals. The proposed method theoretically considers extraction of invariant features to the rotational variations in signals, and the resultant fea-

tures are inherently invariant to the rotations without any pre-processing. The proposed features are based on the frequency information of the signals via FFT and are composed of correlations among all frequency components as well as power spectrum for each frequency. Thus, the proposed method can extract more discriminative features than the standard power spectrum-based methods, while keeping the invariance to the rotations and temporal shift.

2. PROPOSED METHOD

In this section, we describe the proposed method to extract effective features from the three-dimensional acceleration signals. The proposed features are inherently invariant to both three-dimensional rotations and temporal shift.

2.1. Feature extraction

We focus on the frequency information in the three-dimensional sensor signals. The human activity is mainly consisting of cyclic motions, such as walking, running and so on. The characteristics of such cyclic motions are effectively represented by the frequency. We simply employ FFT to extract the frequency information, though wavelet transformation is also applicable for this purpose [5, 7]. Since the FFT can describe the input signals in a frequency domain without any loss, the proposed method is also applicable to extract features from non-cyclic motions as well.

Let $s(t)$ be three-dimensional signals ($s(t) \in \mathbb{R}^3$) at the t-th frame. FFT is applied to the sequence of the signals along respective dimensions:

$$f(\omega) = \int \exp(-i\omega t)s(t)dt \in \mathbb{C}^3, \quad (1)$$

$$F = [f(\omega_1), \cdots, f(\omega_n)] \in \mathbb{C}^{3 \times n}, \quad (2)$$

where n indicates the number of frequencies. The input signals s are transformed into a frequency domain f and then form the complex matrix F whose dimensionality is $3 \times n$.

Next, we consider the auto-correlation matrix of the complex Fourier features F as follows:

$$R = F^*F \in \mathbb{C}^{n \times n}, \quad (3)$$

where F^* denotes complex conjugate transpose of the matrix F. The diagonal elements in R corresponds to the ordinary power spectrum which are summed up over the three-dimensional axes (Fig. 2):

$$R_{jj} = f(\omega_j)^*f(\omega_j) = \sum_{d \in \{x,y,z\}} |f_d(\omega_j)|^2 \in \mathbb{R}, \quad (4)$$

where $|\cdot|$ denotes the absolute value of complex and d indicates the dimension of the signals ($d \in \{x, y, z\}$). These diagonal elements are all real values. On the other hand, the off-diagonal elements are cross-correlations among all frequency
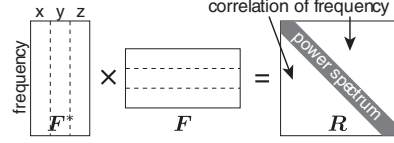

Fig. 2. Auto-correlation matrix R of Fourier features F. The Fourier (complex) features from the three-dimensional signals are multiplied by themselves and then the symmetric auto-correlation matrix R is obtained. The diagonal components of the matrix corresponds to the ordinary power spectrum for each frequency, while the off-diagonal ones are the cross-correlation features among all frequency components.

components (Fig. 2):

$$R_{jk} = f(\omega_j)^*f(\omega_k) = \sum_d f_d(\omega_j)^*f_d(\omega_k) \in \mathbb{C}, \quad (5)$$

and these elements are complex values. By exploiting these correlation features, the relationships among frequencies can be effectively extracted.

Since the auto-correlation matrix R is symmetric, we use only the upper triangle components of which number is $n(n+1)/2$ including the diagonal, and consider their absolute values. The features are actually unfolded to a vector form:

$$z = \text{vector}(\text{upper}(\underline{R})) \in \mathbb{R}^{\frac{n(n+1)}{2} \times 1}, \quad (6)$$

where $\underline{R}$ indicates the matrix consisting of the absolute component values $|R_{ij}|$.

2.2. Invariance

In this section, we show the invariance of the proposed features z to both three-dimensional rotations and temporal shift (delays) of the signals.

If the accelerometer is rotated, the input signals are subject to the three-dimensional rotation denoted by a rotation matrix A:

$$\hat{s}(t) = As(t). \quad (7)$$

The Fourier features F are also subject to that rotation:

$$\hat{f}(\omega) = \int \exp(-i\omega t)\hat{s}(t)dt$$
$$= A\int \exp(-i\omega t)s(t)dt = Af(\omega), \quad (8)$$
$$\therefore \hat{F} = AF. \quad (9)$$

In such case, the auto-correlation matrix R are invariant to A:

$$\hat{R} = \hat{F}^*\hat{F} = F^*A'AF = F^*F = R, \quad (10)$$

where A' denotes the transpose of the matrix A and we use the unitarity $A'A = I$. Therefore, the resultant features z in Eq.(6) are invariant to any rotational variations.

Next, we show the invariance to the temporal shift (delay)

of the signals. The temporal shift is described by

$$\tilde{\boldsymbol{s}}(t) = \boldsymbol{s}(t - \tau). \tag{11}$$

This results in the phase shift in the frequency domain for the Fourier features:

$$\tilde{\boldsymbol{f}}(\omega) = \int \exp(-i\omega t)\tilde{\boldsymbol{s}}_t dt = \int \exp(-i\omega t)\boldsymbol{s}(t - \tau)dt$$

$$= \int \exp(-i\omega\tau)\exp(-i\omega t)\boldsymbol{s}(t)dt$$

$$= \exp(-i\omega\tau)\boldsymbol{f}(\omega), \tag{12}$$

and the absolute values of the components R_{ij} of the auto-correlation matrix $\boldsymbol{R}$ are

$$|\tilde{R}_{jk}| = \left| \tilde{\boldsymbol{f}}(\omega_j)^* \tilde{\boldsymbol{f}}(\omega_k) \right|$$

$$= |\exp(i\omega_j\tau)\exp(-i\omega_k\tau)\boldsymbol{f}(\omega_j)^*\boldsymbol{f}(\omega_k)|$$

$$= |\boldsymbol{f}(\omega_j)^*\boldsymbol{f}(\omega_k)| = |R_{jk}|. \tag{13}$$

Therefore, the proposed features comprising absolute component values of the complex auto-correlation matrix $\boldsymbol{R}$ have invariance to the temporal shift.

As above, the proposed features are inherently *invariant* to both three-dimensional rotation and temporal shift in contrast to the other methods [5, 6] which employ ad hoc preprocessing to increase robustness to such variations.

2.3. Post-processing

Finally, we consider to take the logarithm of the extracted features as post-processing. In general, the human actions are measured by the accelerometer not directly but through various media, such as clothing. Those media affect the signals and the effects are represented by a transfer function $\boldsymbol{b} \in \mathbb{C}^n$ in the frequency domain:

$$\check{\boldsymbol{F}} = \boldsymbol{F}\text{diag}(\boldsymbol{b}), \tag{14}$$

$$\underline{\check{\boldsymbol{R}}} = \underline{\text{diag}(\boldsymbol{b})^* \boldsymbol{R} \text{diag}(\boldsymbol{b})} = \text{diag}(\underline{\boldsymbol{b}})\underline{\boldsymbol{R}}\text{diag}(\underline{\boldsymbol{b}}), \tag{15}$$

where $\text{diag}(\boldsymbol{b})$ indicates the $n \times n$ matrix whose diagonal elements are $\boldsymbol{b}$. By considering the logarithm of the features, we can linearly separate these effects $\boldsymbol{b}$:

$$\log[\underline{\check{\boldsymbol{R}}}] = \log[\underline{\boldsymbol{R}}] + \log[\underline{\boldsymbol{b}}\boldsymbol{b}'], \tag{16}$$

where $\log[\boldsymbol{A}]$ stands for the matrix comprising the logarithm of the component values $\log(A_{ij})$. In this case, linear methods, such as Fisher discriminant analysis and the other multivariate analyses, could cancel out those biases derived from the transfer function $\boldsymbol{b}$. This is the same motivation as in the cepstrum [8] taking the logarithm of power spectrum of audio signals.

3. EXPERIMENTAL RESULTS

We conducted the experiments on gait identification by using the accelerometer embedded in the cellular phone (iPhone®). We collected the signal data at 33Hz sampling rate from 58 persons who freely walk in daily life with holding the cellular phone in hand. There are 2,331 sequences (4∼106 sequences per person, and about 40 sequences on an average). It is quite a challenging task to identify the persons (gait) based on only the acceleration signals under such practical situations without any control for this experiment.

We compared the proposed method to the other methods which extract features from the three-dimensional acceleration signals with preprocessing to increase rotation robustness: the wavelet-based features [5] and the simple statistics based on the gravity vector [6]. The following experimental protocols are employed in all these methods: Running time window of 70 frames is applied to an input sequence with 35 frames step size, and the features are extracted from the sub-sequence within each time window. Thus, we can extract many features (vectors) from the input sequence, and the sequence is classified (identified) by using those features. In a training phase, the method of Fisher discriminant analysis is applied to the extracted features, and project them to lower-dimensional discriminant space ($\mathbb{R}^{57}$). Each extracted feature is classified into one of 58 person categories based on nearest mean of the category in the discriminant space. By voting the classification results to the person categories, the input sequence is finally classified as the category (person) that obtain the maximum votes [4]. The evaluation is performed in 3-fold cross validation. We measured the classification accuracy for each person and then averaged them across all persons in order to avoid the biases in the number of sequences for persons.

First, we evaluated various settings in the proposed method. The Fourier features are split into AC and DC components; DC components are $\boldsymbol{f}(0)$ and AC are $\boldsymbol{f}(\omega), \omega > 0$ in Eq.(1). The DC component indicates the gravity vector as in [6]. We varied the settings in the proposed method as follows: (i) with / without post-processing (logarithm), (ii) only diagonal components (ordinary power spectrum features) / off-diagonal ones / full components containing both, and (iii) AC / AC+DC (full) components. The performance results are shown in Fig. 3. For diagonal features, the DC components degrade the performance and the logarithm significantly improves it. As stated in Sec.2.3, the logarithm is effectively applied in this linear method (FDA). In contrast, for off-diagonal features, the logarithm makes the performance slightly worse and the DC components contributes to the improvement. In this case, by adding DC components, the correlations between AC and DC ones are incorporated. Roughly speaking, by taking the logarithm of the features, the correlations (multiplications) of various components are broken to just the summations, degrading the discriminative power. When both diagonal and off-diagonal features are used, the performance is improved by adding DC components and by taking the logarithm, as an average of the above results. In addition, based on the above results, we combined the logarithmic AC diagonal components and the AC+DC off-diagonal components. The performance is further improved
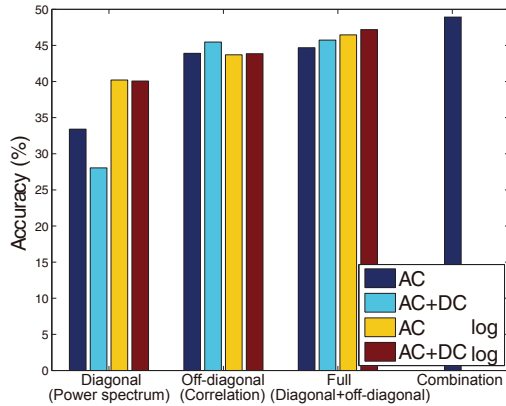
Fig. 3. Performance results for various settings in the proposed method. The feature components are split into diagonal (power spectrum) and off-diagonal ones (correlations), as shown in Fig. 2. The frequencies are also divided into DC and AC components. 'Combination' indicates the method combining the logarithmic AC diagonal components and the AC+DC off-diagonal components.

as shown in the rightmost bar in Fig. 3. As a result, these comparative experiments show that the off-diagonal (correlation) components and the logarithm significantly contribute to the performance improvement, compared to the standard power spectrum features (the leftmost bars in Fig. 3).

Next, we compared the performance of the proposed method to those of the other methods. In the proposed method, the diagonal and off-diagonal components are favorably combined as described above. The results are shown in Fig. 4, demonstrating that the proposed method is superior to the others. The proposed features extracting the correlation among frequencies are quite effective to discriminate the human gaits. It should be noted that the computational cost in the proposed method is also low because the method is almost based on FFT and linear computations without any complicated preprocessing.

4. CONCLUSION

In this paper, we have proposed a method to extract features from three-dimensional acceleration signals. The proposed method computes the correlation matrix of the Fourier transform features which naturally contains both the ordinary power spectrum for each frequency (diagonal) and the correlations among the frequencies (off-diagonal). The proposed features are inherently invariant to rotational variations as well as to temporal shift, although the other methods employ ad hoc preprocessing to increase robustness to those variations. In the experiments on identifying human gaits, the proposed method exhibited the favorable performances compared to the other methods. In this paper, we have focused on the gait identification in the experiments. However, since
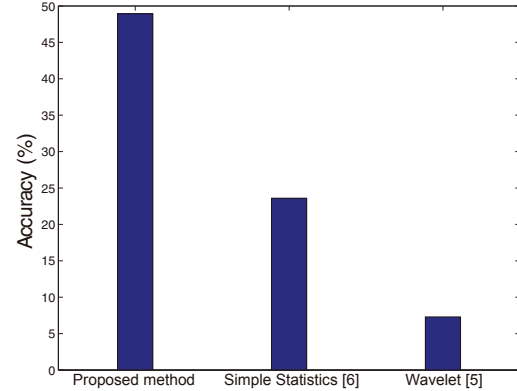

Fig. 4. The performance results of the proposed method compared to the other methods [5, 6]. The proposed method uses the setting of 'combination' (see Fig. 3).

the proposed feature extraction method is general, it can be applicable to the other tasks using the accelerometer, such as segmenting and recognizing various human activities.

5. REFERENCES

[1] S. Sprager and D. Zazula, "Gait identification using cumulants of accelerometer data," in *The 2nd WSEAS International Conference on Sensors, and Signals and Visualization, Imaging and Simulation and Materials Science*, 2009.

[2] J. Mäntyjärvi, M. Lindholm, E. Vildjiounaite, S.-M. Mäkelä, and H. Ailisto, "Identifying users of portable devices from gait pattern with accelerometers," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005.

[3] M.S. Nixon and J.N. Carter, "Automatic recognition by gait," *Proceedings of the IEEE*, vol. 94, pp. 2013–2024, 2006.

[4] T. Kobayashi and N. Otsu, "A three-way auto-correlation based approach to motion recognition," *Pattern Recognition Letters*, vol. 30, no. 3, pp. 185–192, 2009.

[5] J. Mantyjarvi, J. Himberg, and T. Seppanen, "Recognizing human motion with multiple acceleration sensors," in *IEEE International Conference on Systems, Man, and Cybernetics*, 2001, pp. 747–752.

[6] K. Cho, N. Iketani, Y. Okamoto, H. Setoguchi, and M. Hattori, "Implementation and evaluation of human context recognizer for mobile devices with sensors (in japanese)," *Information Processing Society of Japan*, vol. 2010-MBL-52, no. 12, 2010.

[7] T. Iso and K. Yamazaki, "Gait analyzer based on a cell phone with a single three-axis accelerometer," in *The 8th conference on Human-computer interaction with mobile devices and services*, 2006, pp. 141–144.

[8] D. G. Childers, D. P. Skinner, and R. C. Kemerait, "The cepstrum: A guide to processing," *Proceedings of the IEEE*, vol. 65, no. 10, pp. 1428–1443, 1977.