Pattern Recognition 48 (2015) 1289–1301

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Similar gait action recognition using an inertial sensor

Trung Thanh Ngo [a,*], Yasushi Makihara [b], Hajime Nagahara [a,1], Yasuhiro Mukaigawa [c], Yasushi Yagi [b]

[a] Faculty of Information Science and Electrical Engineering, Kyushu University, 744 Motooka, Nishiku, Fukuoka, 819-0395, Japan
[b] The Institute of Scientific and Industrial Research, Osaka University, 8-1Mihogaoka, Ibaraki, Osaka, 567-0047, Japan
[c] Graduate School of Information Science, Nara Institute of Science of Technology, 8916-5 Takayama-cho, Ikoma, Nara, 630-0192, Japan

ARTICLE INFO

Article history:
Received 14 January 2014
Received in revised form
12 August 2014
Accepted 9 October 2014
Available online 22 October 2014

Keywords:
Gait action recognition
Inertial sensor
Heel strike detection
Sensor orientation robustness
Interclass relationship

ABSTRACT

This paper tackles a challenging problem of inertial sensor-based recognition for similar gait action classes (such as walking on flat ground, up/down stairs, and up/down a slope). We solve three drawbacks of existing methods in the case of gait actions: the action signal segmentation, the sensor orientation inconsistency, and the recognition of similar action classes. First, to robustly segment the walking action under drastic changes in various factors such as speed, intensity, style, and sensor orientation of different participants, we rely on the likelihood of heel strike computed employing a scale-space technique. Second, to solve the problem of 3D sensor orientation inconsistency when matching the signals captured at different sensor orientations, we correct the sensor's tilt before applying an orientation-compensative matching algorithm to solve the remaining angle. Third, to accurately classify similar actions, we incorporate the interclass relationship in the feature vector for recognition. In experiments, the proposed algorithms were positively validated with 460 participants (the largest number in the research field), and five similar gait action classes (namely walking on flat ground, up/down stairs, and up/down a slope) captured by three inertial sensors at different positions (center, left, and right) and orientations on the participant's waist.

1. Introduction

With advances in micro-sensor and wireless communication technology, inertial sensors (accelerometer and/or gyroscope) are now low-power, small, accurate, and fast. They are increasingly being embedded in wearable and portable electronic devices such as smartphones, tablets, and smartwatches. As a result, many researchers have been employing a wearable inertial sensor in a variety of research topics such as human-machine interaction [1], user authentication [2], driving analysis [3], fall detection for medical alerts in the elderly [4], rehabilitation and therapy for patients [5], sport training support [6], and a user's daily life surveillance and monitoring [7]. Currently, recognizing a wearer's actions through an inertial sensor is one of the most attractive research topics.

Various actions with different levels of complexity have been investigated in this research field. They are mostly gestures, movements, behaviors, postures, transitions of postures, and sequence of movements of a participant such as sitting, standing, lying, walking, running, walking up/down a slope, falling, driving, cycling, dressing, working in an office, and cooking. Depending on the characteristics of the actions, such as their complexity, periodicity, and dynamicity, the optimal number of sensors and their placement, and recognition method has been decided. We refer readers to a number of recent reports and evaluations [8–14] for details.

There are two essential difficulties for inertial sensor-based action recognition methods: the segmentation of action signals and the relaxation of sensor attachment inconsistency between training and test stages. Particularly, in the case of recognizing similar action classes, an additional difficulty is low recognition accuracy.

Action signal segmentation is the first and most important step toward extracting a signal sequence from an action so that it can be classified. However, existing methods are sensitive to temporal and intensity variation of action signals such as when the participant changes their action speed or style.

The sensor attachment inconsistency problem occurs if locations and/or orientations of the sensor are different between training and testing stages. The existing methods can solve the orientation inconsistency between training and test stages; however, they have to pay a significant loss of signal information. For the details, they have to sacrifice some signal dimension to deal with this problem.

Existing methods have usually been evaluated for relatively different action classes, and hence there is no guarantee that they work well for very similar action classes. Although some authors

* Corresponding author. Tel.: +81 6 6879 8422; fax: +81 6 6877 4375.
E-mail addresses: trungbeo@gmail.com (T.T. Ngo), makihara@am.sanken.osaka-u.ac.jp (Y. Makihara), nagahara@limu.ait.kyushu-u.ac.jp (H. Nagahara), mukaigawa@is.naist.jp (Y. Mukaigawa), yagi@am.sanken.osaka-u.ac.jp (Y. Yagi).
[1] Tel.: +81 92 802 3595; fax: +81 92 802 3579.

http://dx.doi.org/10.1016/j.patcog.2014.10.012
0031-3203/

evaluated their methods with similar action classes such as gait action [15–17], there is no existing method that tentatively solves the problem of similarity of action classes.

In this study, we focus on similar gait action classes, which are the most frequent actions of humans in daily life. We provide solutions to the three above-mentioned problems in the case of classifying similar gait actions:

1. Step signal is detected and segmented employing a scale-space technique. The proposed step detection method can adaptively work with a large amount of variation even if the participant changes their walking speed or style.
2. To solve the practical sensor orientation inconsistency problem. First, we employ a gyroscope for the sensor tilt correction. Then, we apply an orientation-compensative matching algorithm [18] to solve the remaining relative sensor orientation angle between training and test signal sequences. As a result, the proposed method does not experience the information loss problem of existing recognition methods.
3. We propose an algorithm to deal with similar action classes. When action classes are similar, the relationship between one class and all others is more likely to have consistent and distinguished patterns as in the case of gait action. We utilize these relationship patterns to recognize gait action.

This paper is an extended version of our previous work [19]. First, while the previous work did not solve the sensor orientation inconsistency problem, the proposed method does. We employs both an accelerometer and a gyroscope sensors. The advantage of using a gyroscope is that we can fix the sensor tilt (represented by pitch and roll angles) in order to reduce the complexity before applying the orientation-compensative matching algorithm [18] to estimate the remaining orientation angle (yaw). Second, the robustness of step detection against sensor orientation inconsistency is realized and evaluated in this paper. Finally, the previous work evaluated performance using only an attachment location of a single accelerometer of 96 participants. Meanwhile, the proposed method is evaluated rigorously with three variations of sensor orientations and locations and a fourfold increase in the number of participants (460).

2. Related work

2.1. Action signal segmentation

A fixed-size sliding window has frequently been used [20–27]. However, a fixed-size window sometimes introduces errors since it may wrongly segment an action and cannot deal with temporal variation of an action due to speed or user difference. A dynamic window [28,29] has been proposed to solve the problem of the fixed-size window. These methods rely on signal events detected according to a fixed threshold of the signal intensity [28] or noise/signal separation theory to control the size and location of the window. The dynamic windows may, however, still fail when the signal intensity of an action also varies [28]. In the case of gait action recognition, there exist methods [30,31] that detect a gait period (or gait cycle of two consecutive steps) to construct a gait pattern; this is also considered to be using dynamic windows. However, these methods rely on local peak and valley detection, which is sensitive to variations in walking speed and/or style.

2.2. Gait period detection

In the field of inertial gait-based recognition, most existing methods try to detect gait period as a gait primitive, since they work better for the dynamicity of gait signals than those that use a fixed-size sliding window. In such cases, walking is a homogeneous and periodic action, it is hence possible to detect the period of the gait signal by dynamic programming [2], or matching with a sample primitive [32]. However, there is no such method in the field to cope with the situation where gait signal is drastically varied by a number of factors such as intensity, speed, and sensor orientation. The problem is more serious if these factors occur simultaneously.

2.3. Sensor attachment inconsistency

Most existing action recognition methods assume that the sensor is fixed at specific orientation and location on the participant's body. However, it is impractical and unnatural to fix the sensor at the same orientation and location all the time, particularly in daily life (e.g., the sensor orientation of a smartphone in a trouser pocket is subject to change). There are various methods that can be used to solve the sensor location inconsistency (or sensor displacement) problem such as unsupervised adaption [33,34], extracting invariant features from data of different sensor-locations [35], and employing heuristic knowledge [36]. The most popular approach to the sensor orientation inconsistency is to employ a 1D orientation-invariant signal [37,38], which is the magnitude of a 3D signal from an accelerometer or a gyroscope. Other researchers [39–41] use a 2D orientation-invariant signal, which relies on Mizell's research [42], to correct the sensor tilt using a 3D gait acceleration signal. However, these methods produce low performance because of the significant information loss by the dimension reduction of the signal. For the tilt correction, Mizell assumes that the average of 3D acceleration signal samples is the gravity vector in order to correct the sensor tilt. In fact, this assumption does not base on any theory. The averaging of the acceleration samples is performed ignoring the fact that the sensor is rotated when the human body moves. It is particularly incorrect for a short signal sequence that does not contain a natural number of gait periods or when the participant does not walk symmetrically.

A method that corrects the sensor orientation so that all the three dimensions of the signal can be used also exists. However, this method [43] relies on an assumption that the first principal component of the horizontal acceleration data corresponds to the forward (or backward) motion vector. This assumption is not always correct (e.g., when the participant turns), and hence the robustness of the method is reduced. There also exists a method that can estimate 3D relative orientation between a pair of acceleration signal sequences [18]. However this must be carried out for any pair of gallery and probe signal sequences, which is very time-consuming and only suitable for small database problem. In our research, taking the advantage of the gyroscope, we solve the 3D orientation by first estimating the absolute gravity vector to correct the sensor tilt at the pre-processing step and then employing [18] for only solving the remaining relative yaw angle. Consequently, the solution to the sensor orientation inconsistency problem in the proposed method is more advantageous in computational cost, robustness, and accuracy.

Ustev et al. [44] rely on a fusion of sensors to cope with the sensor orientation inconsistency. They use an accelerometer, a compass, and a gyroscope simultaneously to estimate the sensor orientation, hence the captured acceleration signal sequence can be corrected. The limitations of this approach are that the magnetic field is influenced by nearby electronic devices and the signal from the gyroscope is subject to the sensor drift. Moreover, their method needs to know the initial sensor orientation at the beginning of a capturing session of all the participants that limits the application of the method.

2.4. Score normalization

Since the matching scores outputted by the single or multiple matchers are usually heterogeneous, score normalization is needed to transform these scores into a common domain for proper comparison. Score normalization methods, such as Z-normalization [45], T-normalization [46], and cohort-analysis-based normalization [47], find statistical parameters of the score (or similarity) distribution for the transformation model. After a normalization, a new score is computed for each test sample. These score normalization methods are usually used in verification (single matcher case) and in score level fusion (multiple matcher case) [45] for further decisions.

In the proposed method, we use all the matching scores of a test action to all action galleries as an input pattern for the action classifier. This pattern is in fact the interclass relationship pattern of the test action to all the action galleries. Although the magnitude of the pattern is simply normalized, these relationships to all action galleries do not change. This normalization is effective only when the action classes are similar which is the case in this paper. It is used to construct the interclass relationship pattern in an identification problem but not to compare and find the best similarity in a verification problem.

2.5. Interclass relationships

Neeraj et al. [48] proposed a simile classifier for face verification, which uses interclass relationships between a test face and a set of reference faces for recognition with respect to small parts of the face such as the nose, eyes, eyebrows, mouth, and forehead. Our method also employs interclass relationships to improve the recognition performance, but in a much simpler manner. Since all gait action classes are similar, we do not have to tackle the problem of how to choose reference classes and how to divide the feature vectors into smaller feature vectors. In our situation, the whole feature vector is used in the comparison and all gallery action classes are used as reference classes.

3. Assumption and problem setting

Since using one sensor (such as a smartphone) is more practical and natural in real applications than using multiple sensors, and the sensor location about the participant's waist produces good results [43] for these gait actions, we use one inertial sensor that includes a triaxial accelerometer and a triaxial gyroscope.

A participant walks while the sensor is firmly attached to their waist. The orientation of the inertial sensor is therefore fixed in a local coordinate system at the attachment location during each data capturing session for training or testing. On the other hand, sensor orientations in a body coordinate system originated at the body center among participants and sessions may be different.

An action can be recognized using just a gait period, which is considered as the action sample in our problem setting.

The duration of a walking gait period is assumed to be between $T_{min}=700$ ms and $T_{max}=1600$ ms by considering the prior knowledge on natural walking gait styles on large scale and large population databases [49,50].[2] Other gait styles with gait periods outside this range are not considered in this research.

[2] While $[T_{min},T_{max}]$ was found to be [660 ms, 1330 ms] and [740 ms, 1350 ms] in [49,50], respectively, we set it a little bit wider in our implementation.

4. Sensor orientation inconsistency and invariants

4.1. Sensor orientation inconsistency problem

For an inertial sensor fixed on a participant, it can capture the i-th sample of a 3D signal (acceleration or rotational velocity) of the participant's gait at the relative sampling time $t=i\delta$ is described as vector $\boldsymbol{s}_i=(s_{x,i},s_{y,i},s_{z,i})^T$, where δ is the sampling period of the sensor (e.g., $\delta=10$ ms). If another inertial sensor is fixed to the body of the participant at the same location, and the relative sensor orientation between the two sensors is described by a rotation matrix $\boldsymbol{R}'$, the second sensor observes a different signal $\boldsymbol{s}'_i$:

$$\boldsymbol{s}'_i=\boldsymbol{R}'\boldsymbol{s}_i. \tag{1}$$

In the same way, the rotation of a signal sequence $\boldsymbol{S}_u=\langle \boldsymbol{s}_i\rangle (i=1,\ldots,N_S)$ by rotation matrix $\boldsymbol{R}'$ results in another signal sequence $\boldsymbol{S}'_u=\langle \boldsymbol{s}'_i\rangle (i=1,\ldots,N_S)$, where N_S is the number of samples in the sequence, and u stands the acceleration or rotational velocity. This transform is defined as

$$\boldsymbol{R}'\odot\boldsymbol{S}_u:=\langle \boldsymbol{R}'\boldsymbol{s}_i\rangle=\boldsymbol{S}'_u, \tag{2}$$

where $\odot$ is a rotation operator that is applied to a signal sequence.

Therefore, although these inertial sensors capture the same motion, they observe different signals in general and the difference is described by a rotation matrix, which is the relative rotation between two sensor coordinate systems. Thus, in the recognition problem, we cannot directly compare signals captured under different sensor orientations.

4.2. Sensor orientation invariants

Although signal varies due to the sensor orientation change, there are several invariants that help us to deal with the sensor orientation inconsistency.

4.2.1. Magnitude of 3D signal

If we have two inertial sensors whose relative orientation is described by a rotation matrix $\boldsymbol{R}\neq\boldsymbol{I}$, their captured signals $\boldsymbol{s}_i$ and $\boldsymbol{s}'_i$ are different. However, the rotation of a signal does not change its magnitude because $\|\boldsymbol{s}_i\|=\|\boldsymbol{s}'_i\|$. In other words, the magnitude of a 3D inertial signal is invariant to sensor orientation. This invariant information is obviously useful in solving the sensor orientation inconsistency problem. It is called the resultant signal of a 3D inertial signal (3D acceleration or 3D rotational velocity signal).

4.2.2. The earth's gravity

An accelerometer attached to a participant captures both their motion and the earth's gravity. It is difficult to separate the gravity from the captured signal. However, the earth's gravity is a constant vertical acceleration vector in a world coordinate system at the local ground. Although it cannot be used to solve the full 3D sensor orientation, it is still helpful for correcting the sensor tilt as demonstrated in [42].

In the proposed method, we utilize both the invariants in conjunction with a sensor orientation-compensative matching algorithm to solve the full 3D sensor orientation inconsistency problem.

5. Robust step detection

5.1. Gait period

It is well known that a normal gait period consists of a stance phase and a swing phase for each leg [51], the durations of these phases are about 60% and 40% of a gait period, respectively. For a

normal human walk, when the left heel hits the ground in a short moment at the start of the left leg's stance phase, the right foot remains on the ground. The strong impulse of the collision force is transmitted from the left foot to the body center through the left leg, which results in quick motion of the body center. The same phenomenon happens to the motion of the right leg as illustrated in Fig. 1. Therefore, a 3D accelerometer attached at the waist can capture a strong signal at the moment of the *heel strike* (HST) for both legs. Within a gait period, we can observe strong signal vibration at two such moments for the two legs. Beyond the two HSTs, the acceleration signal is smooth and varies gradually, as illustrated in Fig. 2(a). On the other hand, observing the energy consumption of the rectus femoris muscle of a gait period, a clear energy peak can be observed during a HST and it is lower where else [51]. The internal energy consumption can be seen by the externally observed force magnitude at the body center (in form of acceleration magnitude). These characteristics of gait prompt us two statistical observations on signal energy and local peaks/valleys that we use to detect step in the following sections.

Fig. 2 shows an example of a gait period, where the inertial sensor is attached at the center back waist of a participant and its coordinate system coincides with the body coordinate systems so that it can capture up/down, left/right, and backward/forward acceleration as well as the pitch, yaw, and roll of the participant, see Fig. 3. However, if the sensor coordinate system does not coincide with the body coordinate system, a different signal may be observed. This causes the sensor orientation inconsistency problem described in the previous section.

5.2. Step detection based on HST likelihood

Since the gyroscope does not capture HST force, the proposed step detection relies only on the 3D acceleration signal, although a 6D signal is used for recognition.

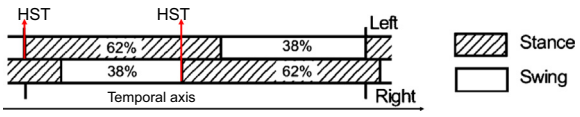

Fig. 1. Time spend on each limb during a gait period of a normal man [51]. The vertical motion of the gait is always influenced by earth gravity, which results in a relatively strong up/down signal.

From the characteristics of gait acceleration described above, we can detect and segment a step based on the characteristics of the HST relying on the computation of its likelihood.

To compute the likelihood of an HST from only the 3D acceleration signal, we rely on two observations on the appearance of an HST:

- *Observation*1: Energy of the acceleration signal is relatively high,
- *Observation*2: The density of local feature points (e.g., peaks and valleys) in all channels is relatively high.

5.2.1. Likelihood based on signal energy

Based on *Observation*1, we regard the energy of the acceleration signal as the likelihood of an HST. Energy $e(i)$ at location $i\delta$ in the time domain is computed as the magnitude of the 3D acceleration signal $e(i) = \|\boldsymbol{s}_{a,i}\|$, which is the orientation-invariant resultant signal. For robustness against temporal variation and noise, we compute different smoothed signal energies $\hat{e}_{\sigma_e,i}$ with Gaussian filters, $f(x,\sigma_e) = \frac{1}{\sqrt{2\pi\sigma_e^2}} e^{-x^2/2\sigma_e^2}$ of different smoothing parameters σ_e. The likelihood of an HST based on the signal energy is

$$p^e(i) = \prod_{\sigma_e} \hat{e}_{\sigma_e,i}. \tag{3}$$

In implementation, since we expect to have two peaks of energy for each gait period with different sizes, we set $\sigma_e \in \{T_{min}/4, T_{min} + T_{max}/8, T_{max}/4\}$, which practically works for different gait periods

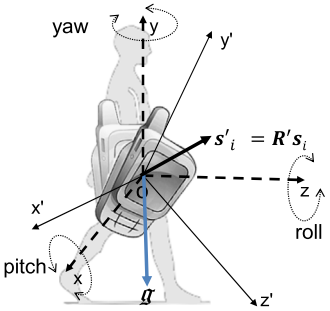

Fig. 3. Initial sensor orientation inconsistency problem. The same human motion may be observed and captured with different signals due to different sensor orientations.

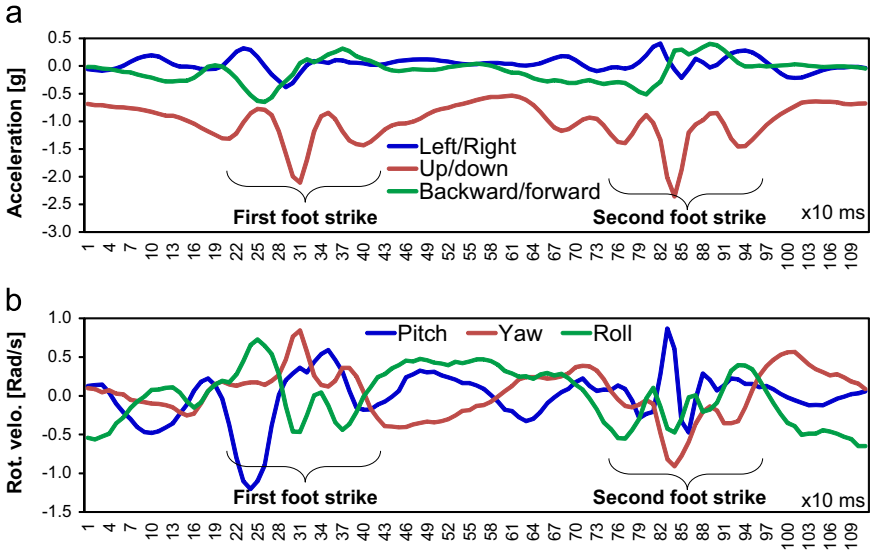

Fig. 2. Example of a 6D inertial signal sequence of a gait period: (a) 3D acceleration and (b) 3D rotational velocity signals. The gait period contains 111 signal samples and its duration is 1110 ms.

from T_{min} to T_{max}. For $\sigma_e < T_{min}/4$, the smoothed energy would contain many peaks, valleys. For $\sigma_e > T_{max}/4$, smoothed energy would be flattened, and hence would not change the quality of the likelihood. Integrating a larger number of over-smoothed energy signal does not improve the quality of energy likelihood, but consumes a higher computational cost. In experiments, we only need one more parameter in the middle of $\{T_{min}/4$ and $T_{max}/4\}$, say $(T_{min}+T_{max})/8$, that works for the medium step duration. Of course, we can use more parameters between $\{T_{min}/4$ and $T_{max}/4\}$, but it is more time-consuming. Fig. 4(b) shows an illustration for different smoothed energies with these values of σ_e.

5.2.2. *Likelihood based on feature density*

Based on *Observation2*, we use the locations of local peaks and valleys for each channel of the signal as the signal features. To get rid of some noise, the signal needs to be smoothed properly. However, it is impossible to get an optimal smoothed signal and the feature detector cannot distinguish HST from non-HST features. There are several phenomena that a feature detector can experience:

- At a fine level of smoothness (signal is weakly smoothed), there are more non-HST features than meaningful HST ones.
- At coarser level, fewer non-HST and more HST features are detected. The total number of detected features decreases.
- At very coarse level (signal is over-smoothed), both HST and non-HST features disappear.

This encourages us to employ a scale-space idea [52,53], in which we consider smoothing the signal at different smoothness levels. When combining the detection results of all smoothness levels, the density of features (regardless whether they are HST or non-HST) is high at the moment of an HST and low otherwise.

The computation of feature density using scale-space idea is illustrated in Fig. 4 as follows. First, the 3D signal sequence, illustrated in Fig. 4(a), is smoothed by several Gaussian filters, $f(x,\sigma_f)$, with different smoothing parameters σ_f. We then detect all the signal features (peaks and valleys) for each channel and each smoothness level, the combined result for each smoothness level is illustrated in Fig. 4(c). Finally, from all the detected feature locations in the time domain $\{i_l\delta\}(l=1,\ldots,N^f)$, where N^f is the number of detected features, the probability density function $p^f(i)$ of features at location i is computed by kernel density estimation as:

$$p^f(i) = \frac{1}{N^f b}\sum_{l=1}^{N^f} K\left(\frac{(i-i_l)\delta}{b}\right), \tag{4}$$

where K is a kernel function used for the estimation, and b is the bandwidth or smoothing parameter used in K. In our implementation we used Epanechnikov kernel [54] and b is set to one half of the minimum duration of a step, $b = T_{min}/4$. The illustration of $p^f(i)$ is shown in Fig. 4(d) for all the features in Fig. 4(c). This probability density function is used as another likelihood of an HST. We use a simple local maximum and minimum within a fixed window of 50 ms in case the sensor's sampling period is 10 ms to detect sharp peak and valley. The smoothing parameter σ_f is initialized with a small value such as 50 ms. A following coarser level is generated by incrementally adding 50 ms to σ_f of its previous level. We can stop smoothing and detecting features when the number of newly detected is less than a threshold (e.g., one feature/sec/signal channel).

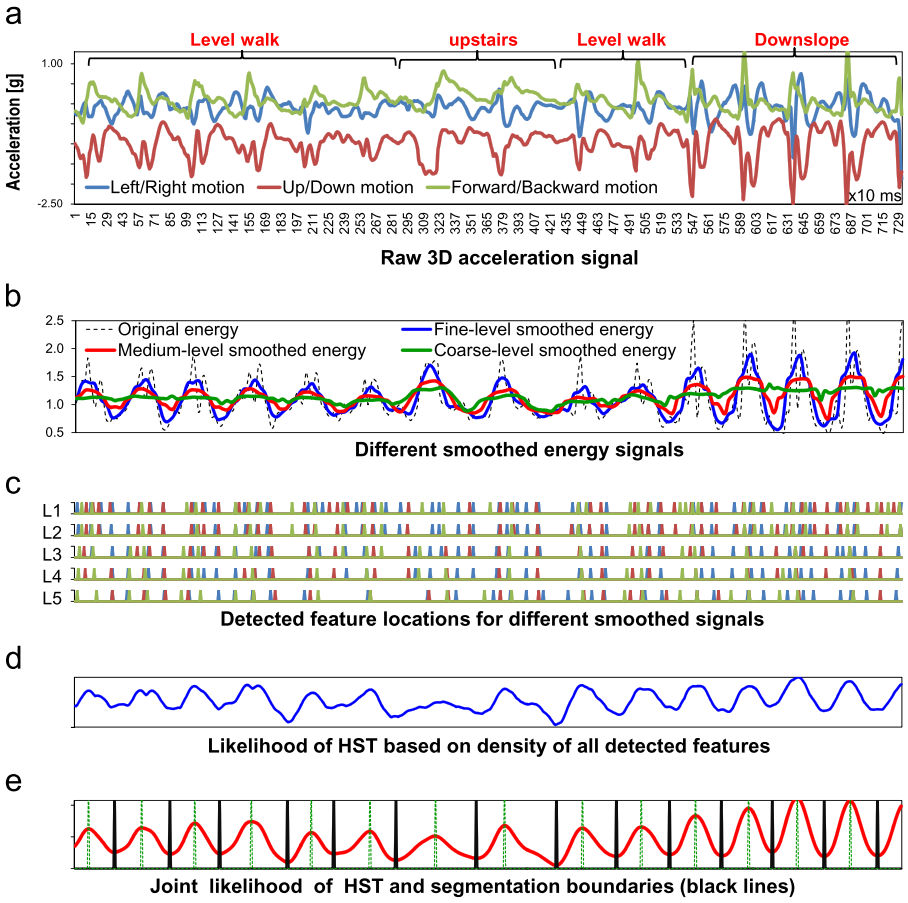

Fig. 4. Illustration of the proposed step detection and segmentation algorithm for a 3D acceleration signal sequence (a) of 5 steps on flat ground, 2 steps upstairs, 2 steps on flat ground, and 4 steps down a slope. All graphs have the same temporal axis as Fig. 2; (a) and (c) use the same color legends. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

5.2.3. Joint likelihood of HST

Considering both *Observation*1 and *Observation*2, the likelihood of HST p_i is computed as the product of two likelihoods $p^e(i)$ and $p^f(i)$:

$$p(i) = p^e(i)p^f(i). \tag{5}$$

Because the HST should contain meaningful information for classifying actions, it would be better to segment the signal into steps so that an HST is located at the center of the segmented step rather than at the segmentation boundaries. The reason is that nonlinear signal matching methods, such as dynamic time warping, would sacrifice some amount of information at the sequence boundaries for matching optimization. A simple local peak detector using a sliding window with size of $T_{min}/2$ can be used to detect HST locations. For illustration, the local peaks of $p(i)$, denoted by dashed green lines in Fig. 4(e), are considered as approximations to the HST locations. A local valley between two adjacent local peaks is used as the segmentation location. Steps are then segmented by all these local valleys, denoted by black lines in Fig. 4(e). Finally, an action sample for recognition is a short signal sequence constructed by two consecutive steps for both acceleration and rotational velocity as illustrated in Fig. 2.

With regard to the robustness to sensor orientation variation, we can obviously compute the feature-based likelihood of an HST relying only on the resultant signal. However, with the resultant signal, we lose several important features as a significant amount of information is lost. Therefore, we also include 3D orientation-dependent signals in computing the feature density-based likelihood to increase the number of HST features at the cost of orientation dependency. However, since the statistical observation on the feature density does not change for any sensor orientation, we will show that this orientation dependency is insignificant in terms of step detection by experiments.

6. Signal matching against sensor orientation inconsistency

To carry out matching between a gallery and test action samples, first, the 6D signals are tilt-corrected, then orientation-compensative matching is carried out with 3D acceleration signals, and finally the distance between 3D rotational velocity signals is computed. The output of this matching algorithm is a pair of distances for acceleration and rotational velocity signals. The flow of the whole matching algorithm is described in Fig. 5.

Before going into the details, we describe the notations for the signals that are used in this section. An action sample of a gait period is described by $\boldsymbol{S} = \langle(\boldsymbol{s}_{a,i}^T, \boldsymbol{s}_{r,i}^T)^T\rangle (i = 1, \ldots, N_S)$, where $\boldsymbol{S}_a = \langle \boldsymbol{s}_{a,i}\rangle$ and $\boldsymbol{S}_r = \langle \boldsymbol{s}_{r,i}\rangle$ are 3D acceleration and rotational velocity signal sequences of $\boldsymbol{S}$, respectively. The subscripts a and r denote the data for acceleration and rotational velocity, respectively. The bold upper-case character (e.g., $\boldsymbol{S}$) stands for a 6D signal sequence; the bold upper-case character with a subscript (e.g., $\boldsymbol{S}_a$ or $\boldsymbol{S}_r$) stands for a 3D signal sequence of acceleration or rotational velocity; and a bold lower-case character (e.g., $\boldsymbol{s}$) stands for a 3D (acceleration or rotational velocity) signal sample.

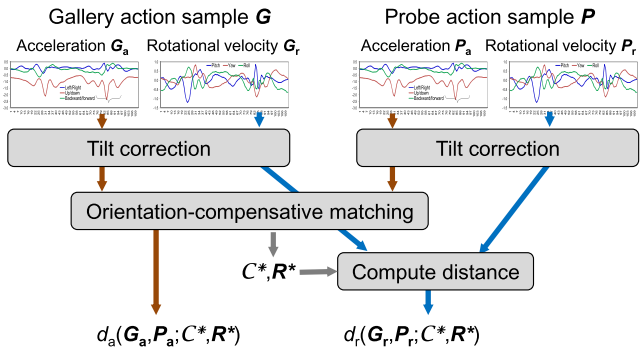

Fig. 5. Signal matching flow. It is noted that tilt correction is performed in advance right after the gait period segmentation.

6.1. Tilt correction at pre-processing

For each 6D action sample of a gait period $\boldsymbol{S} = \langle(\boldsymbol{s}_{a,i}^T, \boldsymbol{s}_{r,i}^T)^T\rangle$ $(i = 1, \ldots, N_S)$, we can compute a rotation matrix $\boldsymbol{R}_i$ of the sensor at the ith frame and ith acceleration signal $\boldsymbol{a}_i$ in a fixed coordinate system f_0 that coincides with the sensor coordinate system at the first frame:

$$\boldsymbol{R}_i = \prod_{j=1}^{i} R(\delta \boldsymbol{s}_{r,j}), \tag{6}$$

$$\boldsymbol{a}_i = \boldsymbol{R}_i \boldsymbol{s}_{a,i}, \tag{7}$$

where $R(\delta \boldsymbol{s}_{r,j})$ is the relative rotation matrix of rotation angles $\delta \boldsymbol{s}_{r,j}$. In a world coordinate system, such as *Oxyz* in Fig. 3, in which gravity $\mathfrak{g} = (0, -1, 0)^T$, the acceleration vector $\boldsymbol{a}_i$ is described by $\boldsymbol{a}_i^w$:

$$\boldsymbol{a}_i^w = \boldsymbol{R}^w \boldsymbol{a}_i, \tag{8}$$

$\boldsymbol{R}^w$ is an unknown constant rotation matrix describing coordinate system transformation. The participant's acceleration $\boldsymbol{a}_i^{w,s}$ can be derived by removing gravity:

$$\boldsymbol{a}_i^{w,s} = \boldsymbol{a}_i^w - \mathfrak{g} = \boldsymbol{R}^w \boldsymbol{a}_i - \mathfrak{g}. \tag{9}$$

The velocity $\boldsymbol{\pi}_i^{w,s}$ of the participant in the world coordinate system is computed by integration:

$$\boldsymbol{\pi}_i^{w,s} = \boldsymbol{\pi}_0 + \sum_{j=1}^{i} (\boldsymbol{R}^w \boldsymbol{a}_j - \mathfrak{g})\delta, \tag{10}$$

where $\boldsymbol{\pi}_0$ is an unknown initial linear velocity in the world coordinate system. If $\boldsymbol{S}$ is an ideal periodic gait period, $\boldsymbol{\pi}_1^{w,s} = \boldsymbol{\pi}_{N_S}^{w,s}$, and $\boldsymbol{R}^w$ is known, then the integration:

$$M_S = \sum_{j=1}^{N_S} (\boldsymbol{R}^w \boldsymbol{a}_j - \mathfrak{g})\delta = 0. \tag{11}$$

Thus, $\boldsymbol{\pi}_0$ defines the direction of participant's motion. For instances, if the participant walks down some stair or a slope, the vector $\boldsymbol{\pi}_0$ points downwards; if the participant walks up a stair of a slope, vector $\boldsymbol{\pi}_0$ points upwards; and if the participant walks on a flat ground, the vector $\boldsymbol{\pi}_0$ is parallel to the ground. Otherwise, if $\boldsymbol{R}^w$ is unknown, we can also find it easily based on the constraint in Eq. (11).

However, in practice, it is difficult to obtain a perfect gait period $\boldsymbol{S}$ (e.g., N_S is shorter or longer than the true value, or the participant does not walk at constant speed), and physically $M_S \neq 0$. We find a solution for $\boldsymbol{R}^w$ by the least squares method:

$$(\boldsymbol{r}^*, \boldsymbol{\pi}_0^*) = \arg\min_{\boldsymbol{r}, \boldsymbol{\pi}_0} \sum_{i=1}^{N_S} \left(\boldsymbol{\pi}_0 + \sum_{j=1}^{i} (R(\boldsymbol{r})\boldsymbol{a}_j - \mathfrak{g})\delta\right)^2 \tag{12}$$

$$\tilde{\boldsymbol{R}}^w = R(\boldsymbol{r}^*), \tag{13}$$

where $\boldsymbol{r}$ is the pitch-yaw-roll vector and $R(\boldsymbol{r})$ is the rotation matrix made from $\boldsymbol{r}$. The minimization is initialized with rotation vector $\boldsymbol{r}_0$ of that $R(\boldsymbol{r}_0)$ forces M_S to be zero. This initialization is different from Mizell's solution [42] in which coordinate system the summation is performed. Mizell sums all acceleration samples without projecting them into the same coordinate system, but we do. Given $\tilde{\boldsymbol{R}}^w$, we can correct for the tilt of the sensor, and then the gait period signals as follows: $\boldsymbol{s}_{a,i}^w = \tilde{\boldsymbol{R}}^w \boldsymbol{s}_{a,i}$, $\boldsymbol{s}_{r,i}^w = \tilde{\boldsymbol{R}}^w \boldsymbol{s}_{r,i}$.

Because there is no reference information in the horizontal plane (perpendicular to the gravity), the estimation of yaw angle

by the minimization is not correct, and therefore any pair of gait periods may differ in their associated yaw angles. In the minimization Eq. (12), we integrate both acceleration and rotational velocity, which may introduce some accumulation error. However, since a gait period is relatively short (about 100 sensor readings if the sensor samples at 100 Hz), the accumulation error is negligible. In this implementation, we use the Levenberg–Marquardt algorithm [55] to solve this minimization.

In the following sections, whenever we mention a gait period, it is assumed to be tilt-corrected at the pre-processing step.

6.2. Orientation-compensative signal matching

In this section, we describe the solution to match a pair of gallery action sample $\boldsymbol{G}=\langle(\boldsymbol{g}_{a,i}^T,\boldsymbol{g}_{r,i}^T)^T\rangle(i=1,\ldots,N_G)$ and probe action sample $\boldsymbol{P}=\langle(\boldsymbol{p}_{a,j}^T,\boldsymbol{p}_{r,j}^T)^T\rangle(j=1,\ldots,N_P)$, where N_G and N_P are the number of signal samples of the action samples $\boldsymbol{G}$ and $\boldsymbol{P}$, respectively. The problem is that these signal sequences are given with different yaw angles, as mentioned above, and hence they cannot be compared directly.

Algorithm 1. Gallery and probe signal registration algorithm.

Require: The gallery and probe signal sequences $\boldsymbol{G}$, $\boldsymbol{P}$
Ensure: Rotation matrix $\boldsymbol{R}^*$ and signal correspondence $\mathcal{C}^*$
$\mathcal{C}^0=DTW(\langle\|\boldsymbol{g}_{a,i}\|\rangle,\langle\|\boldsymbol{p}_{a,j}\|\rangle)$ {Initialization step}
$\gamma^0=\arg\min_\gamma d_a(\boldsymbol{G}_a,\boldsymbol{P}_a;\mathcal{C}^0,R(\gamma))$
$l=0$
repeat
$l=l+1$
$\mathcal{C}^l=DTW(\boldsymbol{G}_a,R(\gamma^{l-1})\odot\boldsymbol{P}_a)$
$\gamma^l=\arg\min_\gamma d_a(\boldsymbol{G}_a,\boldsymbol{P}_a;\mathcal{C}^l,R(\gamma))$
until $\mathcal{C}^l$ and γ^l are converged
$\boldsymbol{R}^*=R(\gamma^l)$
$\mathcal{C}^*=\mathcal{C}^l$

There exists an iterative signal matching method [18], named the orientation-compensative signal matching algorithm, that can efficiently solve the sensor orientation inconsistency problem without reducing signal dimensions. This method simultaneously finds the relative sensor orientation $\boldsymbol{R}^*$ and signal correspondence $\mathcal{C}^*=\{(i_k,j_k)\}(k=1,\ldots,K)$ to minimize the difference between two acceleration signal sequences $\boldsymbol{G}_a=\langle\boldsymbol{g}_{a,i}\rangle$ and $\boldsymbol{P}_a=\langle\boldsymbol{p}_{a,j}\rangle$, where (i_k,j_k) is the kth pair of signal correspondence between the i_kth and j_kth samples of $\boldsymbol{G}_a$ and $\boldsymbol{P}_a$, respectively, and K is the number of correspondence pairs. We then compute the dissimilarity pair (d_a,d_r) between $\boldsymbol{G}$ and $\boldsymbol{P}$ separately for acceleration and rotational velocity using the rotation matrix $\boldsymbol{R}^*$ and signal correspondence $\mathcal{C}^*$ as follows:

$$d_u(\boldsymbol{G}_u,\boldsymbol{P}_u;\mathcal{C}^*,\boldsymbol{R}^*)=\sqrt{\frac{1}{K}\sum_{k=1}^{K}\|\boldsymbol{g}_{u,i_k}-\boldsymbol{R}^*\boldsymbol{p}_{u,j_k}\|^2}, \tag{14}$$

where $u\in\{a,r\}$ stands for acceleration or rotational velocity and $(i_k,j_k)\in\mathcal{C}^*$.

The matching algorithm needs to be carried out for each pair of gallery and probe action samples, which may require a large amount of processing time to match a probe sample with all gallery samples. Since the signals are tilt-corrected, we need to solve only the yaw angle γ difference between these signal sequences, which reduces the computational cost of matching. The matching algorithm is relaxed as summarized in Algorithm 1, where DTW (,,,) is the dynamic time warping between two signal sequences and $R(\gamma)$ is the rotation matrix of the yaw rotation angle γ.

7. Action recognition

7.1. Dissimilarity score to individual action class

A gallery of action templates $\mathbb{G}$ is constructed using 6D action samples generated by training sequences for various participants: $\mathbb{G}=\{\mathcal{G}_i\}(i=1,\ldots,n)$, where $\mathcal{G}_i$ is a collection of action classes i and n is the number of classes. $\mathcal{G}_i$ can be divided into two subsets $\mathcal{G}_{a,i}$ and $\mathcal{G}_{r,i}$ for acceleration and rotational velocity, respectively.

Given a test action sample $\boldsymbol{P}$, we compute a pair of dissimilarity $(D_a(\boldsymbol{P}_a,\mathcal{G}_{a,i}),D_r(\boldsymbol{P}_r,\mathcal{G}_{r,i}))$ between $\boldsymbol{P}$ and gallery action class $\mathcal{G}_i$ for acceleration and rotational velocity, respectively. These dissimilarities can be computed by considering the m smallest dissimilarities between $\boldsymbol{P}$ and the individual gallery action template $\boldsymbol{G}\in\mathcal{G}_i$:

$$D_u(\boldsymbol{P}_u,\mathcal{G}_{u,i})=\frac{1}{m}\sum_{\boldsymbol{G}_u\in NN_u(\boldsymbol{P}_u,\mathcal{G}_{u,i};m)}d_u(\boldsymbol{G}_u,\boldsymbol{P}_u;\mathcal{C}^*,\boldsymbol{R}^*) \tag{15}$$

where $u\in\{a,r\}$, $NN_u(\boldsymbol{P}_u,\mathcal{G}_{u,i};m)$ is a set of m (e.g., 10 in our experiment) nearest neighbors of $\boldsymbol{P}_u$ in $\mathcal{G}_{u,i}$, and $d_u(\boldsymbol{G}_u,\boldsymbol{P}_u;\mathcal{C}^*,\boldsymbol{R}^*)$ is computed by Eq. (14).

7.2. Recognition using interclass relationship

7.2.1. Feature vector

Conventional action recognition approaches usually classify the action by selecting the minimum dissimilarity to template action classes. However, when we try discriminating the similar gait action classes, such minimum criteria are weak at classifying the action.

In addition, we note that not only the dissimilarity to a target action class but also those for the other action classes may contain discriminative information. Therefore, we describe a test sample $\boldsymbol{P}$ by a feature vector composed of normalized dissimilarities to all the gallery action classes separately for acceleration and rotational velocity, $\boldsymbol{v}_{\boldsymbol{P}}=(v_{a,1},\ldots,v_{a,n},v_{r,1},\ldots,v_{r,n})^T$ such that:

$$v_{u,i}=\frac{D_u(\boldsymbol{P}_u,\mathcal{G}_{u,i})}{\sqrt{\sum_{j=1}^{n}D_u(\boldsymbol{P}_u,\mathcal{G}_{u,j})^2}}, \tag{16}$$

where $u\in\{a,r\}$, $i=1,\ldots,n$.

Similarly, each template action sample $\boldsymbol{G}\in\mathcal{G}_i$ is also described by a feature vector $\boldsymbol{v}_{\boldsymbol{G}}$ in a leave-one-out manner. In other words, $\boldsymbol{v}_{\boldsymbol{G}}$ is computed when $\boldsymbol{G}$ is excluded from $\mathbb{G}$. An example of a feature vector that is generated from the level walk sample described in Fig. 2 is illustrated in Fig. 6, in our experiments.

7.2.2. Recognition

Once a training data set $\mathbb{V}=\{\mathcal{V}_i\}(i=1,\ldots,n)$ of $2n$-dimensional feature vector is prepared, where $\mathcal{V}_i$ consists of feature vectors of $\mathcal{G}_i$, a classifier such as SVM or kNN is constructed, and a test action sample $\boldsymbol{P}$ associated with feature vector $\boldsymbol{v}_{\boldsymbol{P}}$ is then classified.

The summary of the proposed method is shown in Fig. 7.

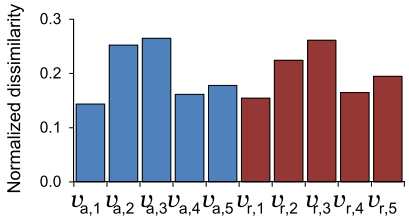

Fig. 6. Example of a generated feature vector for the action sample in Fig. 2.

8. Experiments

8.1. Experimental setup

In our experiments, 3 IMUZ sensors [56] were fixed at the back, left, and right waist of a participant and captured data at a sampling period of 10 ms. Sensors were mounted on a waist belt as shown in Fig. 8(a). The belt was covered by a soft cushion to protect the sensors and avoid direct contact with the participant. When attached, the sensor orientations of the left and right sensors were set at approximately 90° away from the center IMUZ and sensor orientations between the left and right sensors were about 180° (see Fig. 8(c)). In our case, we note that the largest orientation difference corresponded with the largest distance between the sensors. Each participant was asked to walk straight on flat ground, up stairs, down a slope, up a slope, down stairs, and walk straight out of the same environment as shown in Fig. 8(d).

8.2. Datasets and ground-truth

We collected data from 460 participants aged between 8 and 78, the gender ratio was almost equal. All of the data for level walk, and up/downslope walks are published in [57]. We set up two datasets in our experiments. The first dataset contained the whole database captured by three sensors for 460 participants, the dataset was divided randomly into two subsets containing 231 and 229 participants to make training and test action samples, respectively. The details of the age distribution for training and test data are shown in Fig. 9. Since our simulation experiments for the sensor orientation inconsistency problem required a very large number of randomly simulated sensor orientations, we created the second dataset, which is a small subset of the first dataset, containing 125 participants (66 for training and 59 for testing) and captured by the center back IMUZ.

Ground-truth action labels for the signal sequence from the center IMUZ were assigned manually by synchronizing with simultaneously captured videos. Since three IMUZ sensors were easily synchronized, the action labels for signal sequences of the left and right IMUZs were also prepared.

8.3. Benchmark and reference methods

We compared the proposed method (denoted as PROPOSED) with four of the latest benchmark methods, which are summarized in Table 1. The first benchmark method (BOF2012) [26] applies the well-known bag-of-features model to represent the inertial signal by a newly coded string for recognition. The second benchmark method (SIIRTOLA2012) [37] uses the resultant signals from the accelerometer and gyroscope as the orientation-invariant signals to overcome the orientation inconsistency problem. Acknowledging that a significant amount of information is lost using the 1D resultant signal of a 3D acceleration or rotational velocity signal, the third benchmark method (APIWAT2011) [43] corrects the sensor orientation at pre-processing. The fourth benchmark method (NGO2012) [19] is the initial version of the proposed algorithm; the only difference is that it does not solve the sensor orientation inconsistency problem.

We also made some references that were variants of PROPOSED for analysis, as listed in Table 1. INVAR2D, INVAR4D, and INVAR6D use the same step detection and recognition using interclass relationships for recognition, but do not use the orientation compensative matching algorithm. However, these reference methods use orientation-invariant signals (INVAR2D, INVAR4D) or orientation-corrected signals (INVAR6D) to overcome the orientation inconsistency problem. INVAR2D uses a 1D resultant signal from a 3D acceleration or rotational velocity signal that is used by a number of research works [37,38], and thus 2D orientation-invariant signal is used instead of the original 6D signal. Meanwhile, INVAR4D first corrects the vertical motion of an inertial sensor based on Mizell's research [42], where horizontal motion is described by the magnitude of the horizontal signals. Hence, a 2D (1D for vertical and 1D for horizontal motion) orientation-invariant signal is used instead of each 3D signal (acceleration or rotational velocity signal). This 2D orientation-invariant signal is presented in [39–41]. As a result, a 4D orientation-invariant signal is used instead of the original 6D signal. In contrast, INVAR6D corrects the full 6D orientation-invariant signal, similar to APIWAT2011 [43], at pre-processing.

On the other hand, FUS_PRODUCT and FUS_SUM directly use the dissimilarities of the individual action classes without taking the advantage of the interclass relationship for recognition. They use a score-level fusion technique to integrate dissimilarities from the accelerometer and gyroscope. FUS_PRODUCT uses the product rule [58], while FUS_SUM uses the sum rule [58]. Z-normalization [45] is applied before the fusion in these two methods.

For signal segmentation, several parameters of BOF2012 were tuned: primitive size, sample size, and vocabulary size (the number of primitives). The maximum sample size was limited to 200 ms,

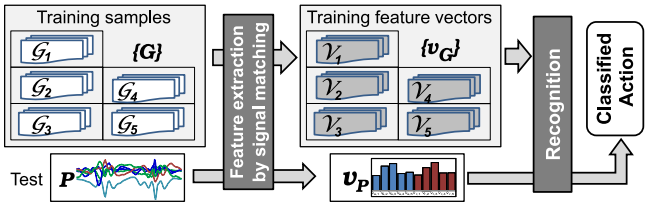

Fig. 7. Flowchart of the proposed recognition algorithm.

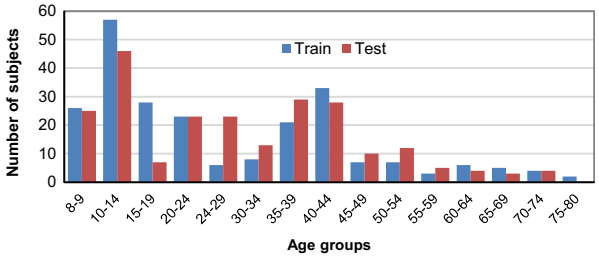

Fig. 9. Histogram of participants by age for training and test subsets of the first dataset.

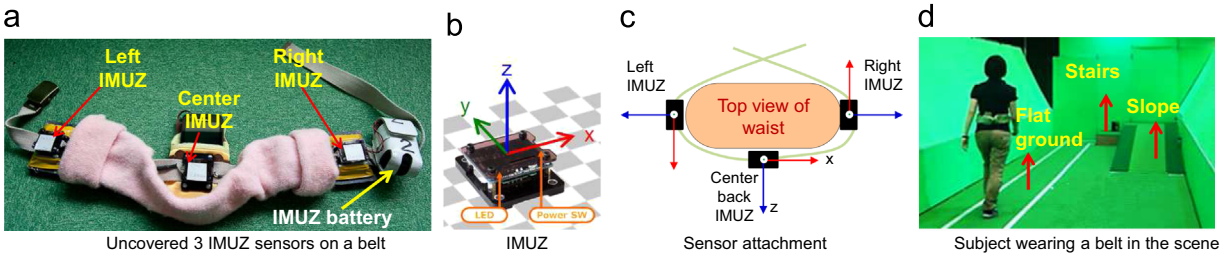

Fig. 8. Experimental setup.

Table 1
Summary of benchmark and reference methods.

Denotation	Signal segmentation	Interclass relationship?	Robust to sensor orientation?	Signal dimension & additional description
BOF2012 [26]	Fixed-size window by exhaustive search	No	No	6D
SIIRTOLA2012 [37]	Fixed-size window of 1 second	No	Yes	2D orientation-invariant signal
APIWAT2011 [43]	Fixed-size window of 1 second	No	Yes	6D orientation-corrected signal
NGO2012 [19]	Robust step detection	Yes	No	6D
PROPOSED	Robust step detection	Yes	Yes	6D
INVAR2D	Robust step detection	Yes	Yes	A variant of the proposed method, 2D sensor orientation invariant resultant signals, same as SIIRTOLA2012, are used
INVAR4D	Robust step detection	Yes	Yes	A variant of the proposed method, 4D sensor orientation invariant signals are used
INVAR6D	Robust step detection	Yes	Yes	A variant of the proposed method, orientation of inertial sensor is corrected, same as APIWAT2011, at the pre-processing, 6D
FUS_PRODUCT	Robust step detection	No	Yes	A variant of the proposed method, product rule fusion of accelerometer and gyroscope is used instead of interclass information, 6D
FUS_SUM	Robust step detection	No	Yes	A variant of the proposed method, sum rule fusion of accelerometer and gyroscope is used instead of interclass information, 6D

which is assumed to be the upper limit for a normal human walking cycle. We carried out an exhaustive search to find the best parameters for BOF2012: a primitive size of 5 ms, vocabulary size of 14, and sample size of 200 ms. For APIWAT2011 and SIIRTOLA2012, the fixed-size window was 1 s, which is the same as in their original experiments. For the machine learning technique in APIWAT2011, SIIRTOLA2012, and all variants of the proposed method, SVMLib [59] was selected with the option of multiple binary classifiers using a linear kernel.

8.4. Results and discussion

8.4.1. Evaluation of step detection

First, we performed the proposed step detection to find a segmentation boundary set $\{l\}$ (illustrated by black lines in Fig. 4(e)) for the action signal sequence of the center back IMUZ of the first dataset. Then, we manually checked the results relying on the actual signal intensity and captured videos. The proposed algorithm worked perfectly on the action signal sequences from the center back IMUZ. In any experiment in this section, we first detect a new set of segmentation boundaries $\{l'\}$, then the average difference compared with the baseline $\{l\}$ was used to evaluate the segmentation performance.

In the first experiment, we carried out a simulation experiment for step detection with the center IMUZ of the first dataset. For each trial, a random 3D rotation using Rodrigues' rotation formula was generated. The step detection performance was checked under various simulated sensor orientations. The average result of 10 random trials for each rotation magnitude is shown in Fig. 10. At zero sensor orientation, the step detection performance was the same as the baseline. The difference occurred if the sensor orientation differed from the original configuration as we also include the original (orientation-dependent) signal in the detection, which is mentioned at the end of Section 5. However, the average step location difference was just about 2 ms, which is insignificant since the sampling period was 10 ms and the walking gait period was about 1000 ms.

In the second experiment, the step detection performances was measured from the left and right IMUZs of the first dataset. Since the three sensors were easily synchronized, the step detection performance on the center IMUZ was used to evaluate those on the left and right ones similar to the first experiment. In our sensor setup, the left and right IMUZs were set at an orientation of about 90° away from the center IMUZ and the distance between the left or right and the center sensor differentiated the acceleration signals captured across them. However, the average difference was less than 3 ms as shown in Table 2.

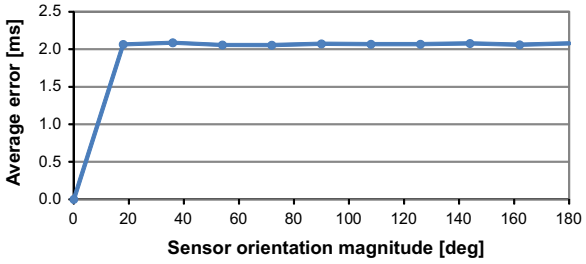

Fig. 10. Performance of step detection against simulated sensor orientation.

Table 2
Average step location differences from the baseline for left and right sensors.

Diff. (ms)	Left IMUZ	Right IMUZ
Mean	2.71	2.75
Std. Dev.	1.11	1.41

In addition, we made a histogram of actual detected periods, shown in Fig. 11 for the center IMUZ of the first dataset. We can see that the existing parameters of Okumura2010 [49] ([660 ms, 1330 ms]) and Oberg93 [50] ([740 ms, 1350 ms]) with narrower ranges also fit our database. Therefore, it is unlikely that we need to change our parameters for a new walking database. However, we made an additional experiment trying different parameters for the same dataset to evaluate the scale-space technique in our step segmentation, which is summarized in Table 3. We borrowed the parameters of Okumura2010, Oberg93, and a new, wider setting:

[600 ms, 1600 ms]. Other parameters such that $T_{min} < 600$ ms and $T_{max} > 1600$ ms are out of scope. We can clearly see that the difference between these segmentation results and the baseline (with our parameter setting) is insignificant. This also implies that the proposed period segmentation can be applied for broader range of gait activity including jogging, running, and sprinting with slightly shorter gait period.

Overall, from the results in this section, we can see that the proposed step detection is robust and accurate (manually checked). The reason for the robustness to sensor orientation inconsistency is that we rely on statistics that do not depend on the sensor orientation: the magnitude and feature density of the acceleration signal. The reason for the robustness to parameter setting is explained by the merit of scale-space technique.

8.4.2. Segmented gait action samples and interclass relationship feature vectors

Examples of segmented action samples are shown in Fig. 12(a) for five action classes of the whole training set. In Fig. 12(b), the distribution of feature vectors for each action class is described by a mean vector and standard deviations that are illustrated by a bar graph with error bars. We see that the gait action periods are well segmented, and that the interclass relationships have clear and relatively distinguished patterns for each action class, which strongly encourages the use of the proposed recognition algorithm.

For example, upslope patterns sometimes appear very similar to level walk patterns. Given an upslope sample, its dissimilarities to the level walk and the upslope classes are therefore very similar (see Fig. 12(b_4), the first and the fourth bar ($v_{a,1}$ and $v_{a,4}$) for acceleration, and the sixth and the ninth bar ($v_{r,1}$ and $v_{r,4}$) for rotational velocity), and we may often mis-classify the upslope sample when we use only the individual dissimilarities. On the other hand, when we consider the level walk sample and the upslope sample from the viewpoint of the interclass relationship feature vector, we notice that the upslope sample produces larger dissimilarities to the downslope class than to the level walk class (see Fig. 10 (b_1) and (b_4), the fifth and tenth bars for acceleration and angular velocity, respectively). This implies that the interclass relationship pattern helps us to more accurately classify the similar gait action classes than individual dissimilarities.

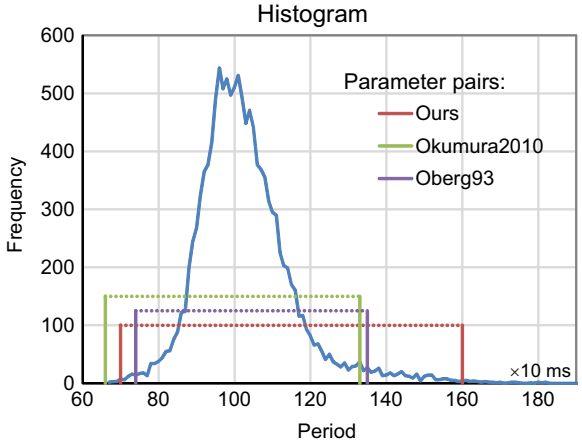

Fig. 11. Histogram of detected periods.

Table 3
Average step location differences from the baseline for parameter settings.

Diff. (ms)	Okumura2010	Oberg93	[600, 1600]
Mean	2.18	1.96	2.36
Std. Dev.	1.16	1.12	1.21

8.4.3. Recognition experiment against simulated sensor orientation

In the first recognition experiment, we evaluated the proposed method against various simulated sensor orientation differences between training and test signals from the second dataset. Signals from the test dataset were rotated by random 3D rotation vector using Rodrigues' rotation representation to simulate the sensor orientation inconsistency. We compared the performance of PROPOSED with those of the benchmark and references methods. The average results of 10 random trials for each magnitude of sensor orientation are described in Fig. 13.

We can clearly see that BOF2012 and NGO2012 could only work when the sensor orientation difference between training and test data was small. This is also when orientation-compensative matching in PROPOSED is considered unnecessary. However, the performance of NGO2012 is worse than that of PROPOSED. The fact is that although we planned to fix the IMUZ at the same orientation, it was difficult to do so for all the participants as their body tilt could unexpectedly change while they were walking. That resulted in a slight sensor orientation inconsistency problem. That is why PROPOSED, which is equipped with the orientation-compensative matching algorithm, produced a slightly better result compared with that of NGO2012. We also can see that for larger orientation differences, BOF2012 and NGO2012 gave worse accuracy. This problem arose because they

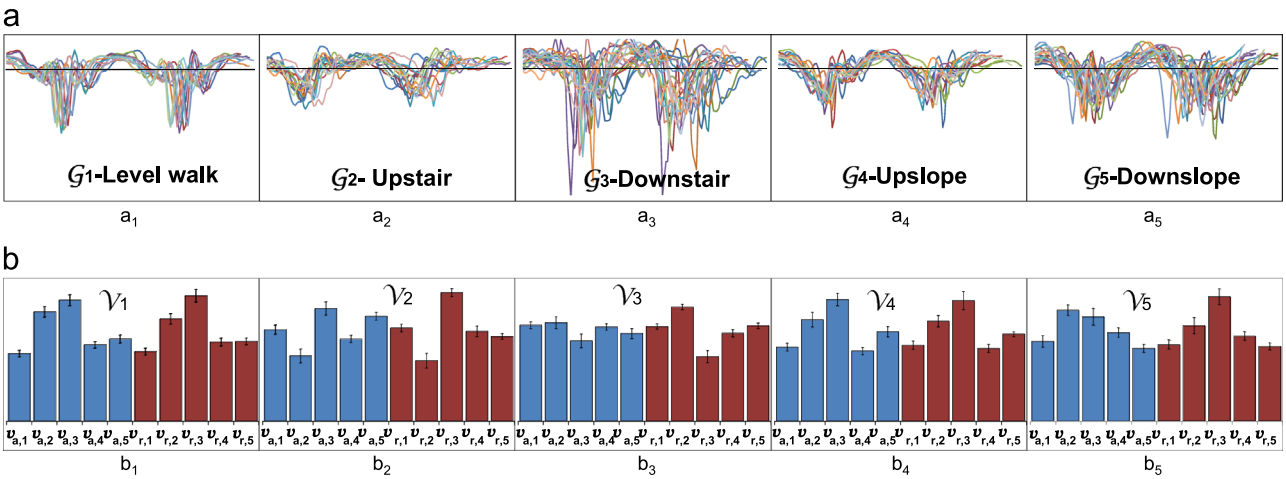

Fig. 12. Segmented gait action samples of training feature vectors: (a) raw action samples of gait periods on the up/down acceleration signal of 5 classes, and (b) the distributions of their represented feature vectors.

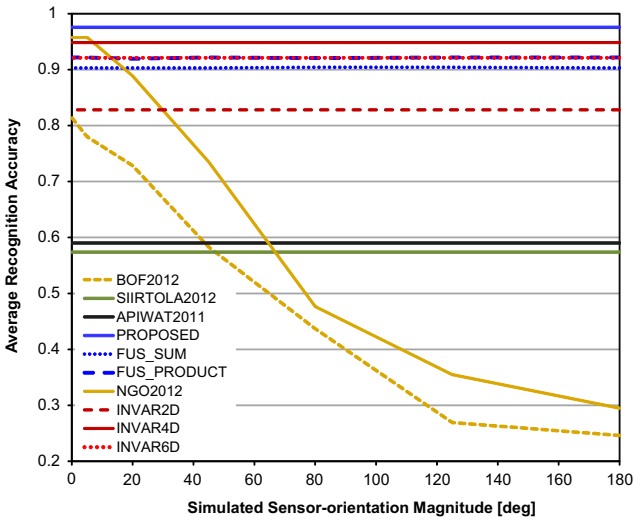

Fig. 13. Recognition performances of proposed, benchmark, and reference methods against simulated sensor orientation.

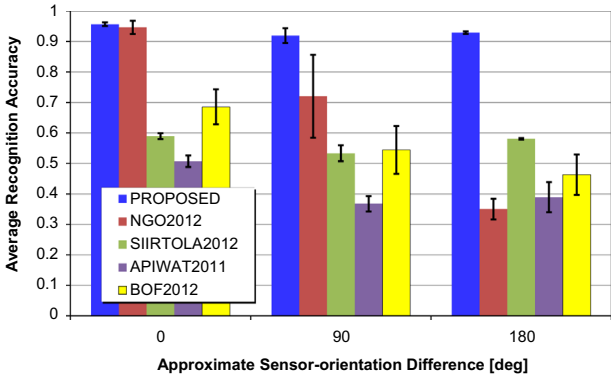

Fig. 14. Statistical results of the performances of benchmark methods considering all combinations of three sensors.

did not consider the sensor orientation inconsistency problem. Meanwhile, all other reference and benchmark methods were robust against the simulated sensor orientation, their performance remained similar for any sensor orientation difference between training and test data.

Although SIIRTOLA2012 and APIWAT2011 were robust against the sensor orientation inconsistency problem, their recognition accuracies were relatively low due to the limitation of the fixed-size window.

With regard to approaches to generating sensor orientation invariant signal, INVAR2D was inferior to INVAR4D due to larger amount of information loss (1D resultant signal compared with 2D resultant signal for acceleration and rotational velocity separately, see Section 8.3). INVAR6D was inferior to INVAR4D because additional horizontal 2D signals are not compensated correctly in theory as mentioned in Section 2.3. PROPOSED outperformed these references, because the proposed method treats full 6D signal whose orientation is correctly compensated.

With regard to the usage of interclass relationships, we can see that PROPOSED with interclass relationships outperformed FUS_SUM and FUS_PRODUCT without interclass relationships for recognition.

Overall, PROPOSED gave the best accuracy and robust performance against various sensor orientations.

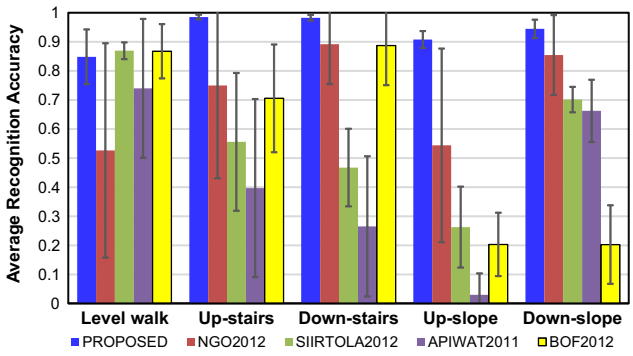

Fig. 15. Average accuracies for 5 action classes. The legends are the same with those of Fig. 14.

Table 4
Average confusion matrix (%) of proposed method.

Test action	Predicted action				
	LW	UT	DT	UL	DL
Level walk (LW)	84.76	0.00	0.00	11.36	3.88
Upstairs (UT)	0.05	98.44	0.00	1.46	0.05
DownStairs (DT)	0.00	0.00	98.22	1.09	0.69
Upslope (UL)	7.36	0.02	0.10	90.84	1.67
Downslope (DL)	4.43	0.02	0.02	1.01	94.53

8.4.4. Overall experiment

In this section, we carried out an overall experiment on the first dataset. For each IMUZ sensor, we had a training (denoted as a single capital character T) subset and a test (denoted as E) subset. Considering 3 training and 3 test subsets of left (L), right (R), and center (C) IMUZ sensors, we had 9 combinations in total and they could be categorized in term of sensor orientation difference between a training and a test subsets: 0-degree-difference combinations {(LT,LE), (CT,CE), (RT,RE)}; 90-degree-difference combinations {(LT,CE), (RT,CE), (CT,LE), (CT,RE)}; and 180-degree-difference combinations {(LT,RE), (RT,LE)}.

The average accuracies of all the action classes and their average are shown in Fig. 14 with a standard deviation error bar for each method. We can see a similar trend to that shown in Fig. 13. The performance of SIIRTOLA2012 was robust to sensor orientation inconsistency while the performances of NGO2012 and BOF2012 decreased as the sensor orientation difference became larger. The performance of NGO2012 at the same sensor configuration (0-degree-difference) was also slightly worse than that of PROPOSED for the same reasons as explained in the previous experiment, Section 8.4.3. The performance of APIWAT2011 differed from that in Fig. 13. The reason is explained as follows. Although APIWAT2011 applied the sensor orientation correction, its assumptions were not correct for the sensors attached at different locations on waist. It is because, since accelerations at different locations on the same participant are mechanically different, so are the statistical properties. That is why its performance was not as robust as we have seen in Fig. 13, where the sensor was at the same location. In contrast, the performance of PROPOSED was accurate and robust for any sensor orientation difference.

We also prepared the average accuracies of all combinations for each action class as shown in Fig. 15. The performances of BOF2012 and SIIRTOLA2012 were high for the level walk action but low for the other action classes. A large variation in performance was also seen in NGO2012 and APIWAT2011. In contrast, PROPOSED accurately and stably worked for all five action classes.

The average confusion matrix for PROPOSED, used in Fig. 15, is shown in Table 4 for the details on the recognition performances of the five actions. From the table, upslope and downslope walks

are sometime confused with level walk. Meanwhile, upstairs and downstairs walks are the two most distinguishable among the five gait actions. The results can partially be seen by checking the magnitude of the vertical acceleration illustrated in Fig. 12.

From this experiment, we also can see that the proposed method can practically overcome some amount of sensor displacement around the participant's waist.

9. Conclusions

We proposed a recognition method for similar gait actions represented by signal sequence as short as a gait period (about 1 s) using an inertial sensor. First, we proposed a robust step detection method based on scale-space technique to segment a signal into action samples. The method is designed to work well even if the action drastically varies in speed or intensity. Second, we presented a solution to deal with the sensor orientation inconsistency problem. Third, we also proposed a recognition method using interclass relationships to overcome the problem of similar action classes. Experiments for five similar gait action classes (walking on flat ground, up stairs, down stairs, up a slope, and down a slope) of a very large number of participants (460) positively validated the proposed method, while the existing methods have been evaluated with less than a hundred participants.

Although the proposed method is designed to overcome the sensor orientation inconsistency, it has the practical possibility to work against some amount of sensor location inconsistency.

In future, we would like to extend the number of gait action classes for real application and find out if the recognition method is useful for other data in which the sample classes are very similar.

Conflict of interest

None.

Acknowledgments

This work was supported by JSPS Grant-in-Aid for Scientific Research(S) 21220003.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.patcog.2014.10.012.

References

[1] R. Raya, E. Rocon, J.A. Gallego, R. Ceres, J.L. Pons, A robust kalman algorithm to facilitate human-computer interaction for people with cerebral palsy, using a new interface based on inertial sensors, Sensors 12 (3) (2012) 3049–3067.
[2] T.T. Ngo, Y. Makihara, H. Nagahara, R. Sagawa, Y. Mukaigawa, Y. Yagi, Phase registration in a gallery improving gait authentication, in: Proceedings of the International Joint Conference on Biometrics, 2011, pp. 1–7.
[3] J. Paefgen, F. Kehr, Y. Zhai, F. Michahelles, Driving behavior analysis with smartphones: insights from a controlled field study, in: Proceedings of the Eleventh International Conference on Mobile and Ubiquitous Multimedia, MUM '12, ACM, New York, NY, USA, 2012, pp. 36:1–36:8.
[4] N. Noury, A. Fleury, P. Rumeau, A. Bourke, G. Laighin, V. Rialle, J. Lundy, Fall detection - principles and methods, in: Engineering in Medicine and Biology Society, 2007. EMBS 2007. Twenty-ninth Annual International Conference of the IEEE, 2007, pp. 1663–1666.
[5] C. Cifuentes, A. Braidot, L. Rodriguez, M. Frisoli, A. Santiago, A. Frizera, Development of a wearable ZigBee sensor system for upper limb rehabilitation robotics, in: Fourth IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob), 2012, 2012, pp. 1989–1994.
[6] L. Cheng, S. Hailes, Analysis of wireless inertial sensing for athlete coaching support, in: Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE, 2008, pp. 1–5.
[7] S. Zhang, M. Ang, W. Xiao, C. Tham, Detection of activities for daily life surveillance: Eating and drinking, in: Tenth International Conference on e-health Networking, Applications and Services, 2008. HealthCom 2008, 2008, pp. 171–176.
[8] S. Preece, J. Goulermas, L. Kenney, D. Howard, A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data, IEEE Trans. Biomed. Eng. 56 (3) (2009) 871–879.
[9] S.J. Preece, J.Y. Goulermas, L.P.J. Kenney, D. Howard, K. Meijer, R. Crompton, Activity identification using body-mounted sensors-a review of classification techniques, Physiol. Meas. 30 (4) (2009) R1.
[10] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, P.J.M. Havinga, Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: a survey, in: International Conference on Architecture of Computing Systems, Workshop Proceedings, 2010, pp. 1–10.
[11] K. Altun, B. Barshan, O. Tuncel, Comparative study on classifying human activities with miniature inertial and magnetic sensors, Pattern Recognit. 43 (10) (2010) 3605–3620.
[12] V.H. Cheung, L. Gray, M. Karunanithi, Review of accelerometry for determining daily activity among elderly patients, Arch. Phys. Med. Rehabil. 92 (6) (2011) 998–1014.
[13] L. Chen, J. Hoey, C. Nugent, D. Cook, Z. Yu, Sensor-based activity recognition, IEEE Trans. Syst. Man Cybern. C: Appl. Rev. 42 (6) (2012) 790–808.
[14] Y. Meng, H.-C. Kim, A review of accelerometer-based physical activity measurement, in: K.J. Kim, S.J. Ahn (Eds.), Proceedings of the International Conference on IT Convergence and Security 2011, Lecture Notes in Electrical Engineering, vol. 120, Springer Netherlands, 2012, pp. 223–237.
[15] M. Sekine, T. Tamura, M. Akay, T. Fujimoto, T. Togawa, Y. Fukui, Discrimination of walking patterns using wavelet-based fractal analysis, IEEE Trans. Neural Syst. Rehabil. Eng. 10 (3) (2002) 188–196.
[16] S. Bonnet, P. Jallon, Hidden markov models applied onto gait classification, in: Eighteenth European Signal Processing Conference, EURASIP, 2010, pp. 929–933.
[17] R. Ibrahim, E. Ambikairajah, B. Celler, N. Lovell, Time-frequency based features for classification of walking patterns, in: Fifteenth International Conference on Digital Signal Processing, 2007, 2007, pp. 187–190.
[18] T.T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, Y. Yagi, Orientation-compensative signal registration for owner authentication using an accelerometer, IEICE Trans. Inf. Syst. E97-D (3) (2014) 541–553.
[19] T.T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, Y. Yagi, Inertial-sensor-based walking action recognition using robust step detection and inter-class relationships, in: Twenty-first International Conference on Pattern Recognition, 2012, pp. 3811–3814.
[20] L. Bao, S. Intille, Activity recognition from user-annotated acceleration data, in: A. Ferscha, F. Mattern (Eds.), Pervasive Computing, Lecture Notes in Computer Science, vol. 3001, Springer Berlin Heidelberg, 2004, pp. 1–17.
[21] N. Ravi, N. Dandekar, P. Mysore, M.L. Littman, Activity recognition from accelerometer data, in: Proceedings of the Seventeenth Conference on Innovative Applications of Artificial Intelligence, vol. 3, IAAI'05, AAAI Press, 2005, pp. 1541–1546.
[22] N. Wang, E. Ambikairajah, N. Lovell, B. Celler, Accelerometry based classification of walking patterns using time-frequency analysis, in: International Conference of the IEEE Engineering in Medicine and Biology Society, 2007, pp. 4899–4902.
[23] A. Khan, Y.-K. Lee, S. Lee, T.-S. Kim, A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer, IEEE Trans. Inf. Technol. Biomed. 14 (5) (2010) 1166–1172.
[24] A. Mannini, A. Sabatini, On-line classification of human activity and estimation of walk-run speed from acceleration data using support vector machines, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, 2011, pp. 3302–3305.
[25] J.R. Kwapisz, G.M. Weiss, S.A. Moore, Activity recognition using cell phone accelerometers, SIGKDD Explor. Newsl. 12 (2) (2011) 74–82.
[26] M. Zhang, A.A. Sawchuk, Motion primitive-based human activity recognition using a bag-of-features approach, in: Proceedings of the Second ACM SIGHIT International Health Informatics Symposium, 2012, pp. 631–640.
[27] D. Kelly, B. Caulfield, An investigation into non-invasive physical activity recognition using smartphones, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2012 , 2012, pp. 3340–3343.
[28] J.O. Laguna, A.G. Olaya, D. Borrajo, A dynamic sliding window approach for activity recognition, in: Proceedings of the nineteenth International Conference on User modeling, Adaption, and Personalization, 2011, pp. 219–230.
[29] M. Nyan, F. Tay, K. Seah, Y. Sitoh, Classification of gait patterns in the time-frequency domain, J. Biomech. 39 (14) (2006) 2647–2656.
[30] S.-W. Lee, K. Mase, Activity and location recognition using wearable sensors, IEEE Pervasive Comput. 1 (3) (2002) 24–32.
[31] H. Ying, C. Silex, A. Schnitzer, S. Leonhardt, M. Schiek, Automatic step detection in the accelerometer signal, in: Fourth International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007), IFMBE Proceedings, vol. 13, Springer Berlin Heidelberg, 2007, pp. 80–85.
[32] M.O. Derawi, P. Bours, K. Holien, Improved cycle detection for accelerometer based gait authentication, in: Proceedings of the Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2010, pp. 312–317.
[33] R. Chavarriaga, H. Bayati, J.D. Millán, Unsupervised adaptation for acceleration-based activity recognition: robustness to sensor displacement and rotation, Pers. Ubiquitous Comput. 17 (3) (2013) 479–490.

[34] K. Forster, D. Roggen, G. Troster, Unsupervised classifier self-calibration through repeated context occurences: is there robustness against sensor displacement to gain?, in: International Symposium on Wearable Computers, 2009. ISWC '09, 2009, pp. 77–84.
[35] K. Forster, P. Brem, D. Roggen, G. Troster, Evolving discriminative features robust to sensor displacement for activity recognition in body area sensor networks, in: Fifth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2009, 2009, pp. 43–48.
[36] K. Kunze, P. Lukowicz, Dealing with sensor displacement in motion-based onbody activity recognition systems, in: Proceedings of the Tenth International Conference on Ubiquitous Computing, UbiComp '08, ACM, New York, NY, USA, 2008, pp. 20–29.
[37] P. Siirtola, J. Röning, Recognizing human activities user-independently on smartphones based on accelerometer data, Int. J. Interact. Multim. Artif. Intell. (2012) 38–45.
[38] P. Widhalm, P. Nitsche, N. Braendle, Transport mode detection with realistic smartphone sensor data, in: Twenty-first International Conference on Pattern Recognition, 2012, pp. 573–576.
[39] C.W. Han, S.J. Kang, N.S. Kim, Implementation of HMM-based human activity recognition using single triaxial accelerometer, IEICE Trans. Fundam. Electron. Commun. Comput. Sci. E93-A (2010) 1379–1383.
[40] B. Florentino-Liano, N. O'Mahony, A. Artes-Rodriguez, Human activity recognition using inertial sensors with invariance to sensor orientation, in: Third International Workshop on Cognitive Information Processing (CIP), 2012, 2012, pp. 1–6.
[41] J. Yang, Toward physical activity diary: motion recognition using simple acceleration features with mobile phones, in: Proceedings of the First International Workshop on Interactive Multimedia for Consumer Electronics, ACM, New York, NY, USA, 2009, pp. 1–10.
[42] D. Mizell, Using gravity to estimate accelerometer orientation, in: Proceedings of the Seventh IEEE International Symposium on Wearable Computers, 2003, 2003, pp. 252–253.
[43] A. Henpraserttae, S. Thiemjarus, S. Marukatat, Accurate activity recognition using a mobile phone regardless of device orientation and location, in: International Conference on Body Sensor Networks (BSN), 2011, 2011, pp. 41–46.
[44] Y.E. Ustev, O. Durmaz Incel, C. Ersoy, User, device and orientation independent human activity recognition on mobile phones: Challenges and a proposal, in: Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, UbiComp '13 Adjunct, ACM, 2013, pp. 1427–1436.
[45] A. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, Pattern Recognit. 38 (12) (2005) 2270–2285.
[46] R. Auckenthaler, M. Carey, H. Lloyd-Thomas, Score normalization for text-independent speaker verification systems, Digit. Signal Process. 10 (2000) 42–54.
[47] G. Aggarwal, N. Ratha, R. Bolle, Biometric verification: Looking beyond raw similarity scores, in: Conference on Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06, 2006, pp. 31–31.
[48] N. Kumar, A.C. Berg, P.N. Belhumeur, S.K. Nayar, Attribute and simile classifiers for face verification, in: The Twelfth IEEE International Conference on Computer Vision (ICCV), 2009, pp. 365–372.
[49] M. Okumura, H. Iwama, Y. Makihara, Y. Yagi, Performance evaluation of vision-based gait recognition using a very large-scale gait database, in: IEEE Fourth International Conference on Biometrics: Theory, Applications and Systems, 2010, pp. 1–6.
[50] T. Oberg, A. Karsznia, Basic gait parameters: reference data for normal subjects, 10–79 years of age, J. Rehabil. Res. Dev. 30 (1993) 210–223.
[51] C.L. Vaughan, B.L. Davis, J.C.O. Connor, Dynamics of Human Gait, second ed., Kiboho Publishers, 1999, pp. 7–14 (Chapter 2).
[52] A.P. Witkin, Scale-space filtering, in: Proceedings of the Eighth International Joint Conference on Artificial Intelligence, vol. 2, IJCAI'83, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1983, pp. 1019–1022.
[53] J. Weickert, S. Ishikawa, A. Imiya, Linear scale-space has first been proposed in Japan, J. Math. Imaging Vis. 10 (3) (1999) 237–252.
[54] M. Wand, C. Jones, Kernel Smoothing, Monographs on Statistics and Applied Probability, Chapman & Hall, 1995, pp.10–52 (Chapter 2).
[55] M. Lourakis, levmar: Levenberg-Marquardt nonlinear least squares algorithms in C/C++, [web page] ⟨http://www.ics.forth.gr/~lourakis/levmar/⟩, [Accessed on 31 January 2005.] (2004).
[56] ZMP Inc., IMUZ, ⟨http://www.zmp.co.jp/products/imu-z⟩, In Japanese.
[57] T.T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, Y. Yagi, The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication, Pattern Recognit. 47 (1) (2014) 228–237.
[58] J. Kittler, M. Hatef, R.P. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.
[59] C.-C. Chang, C.-J. Lin, Libsvm: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) 27:1–27:27.

Thanh Trung Ngo received his M.E and Ph.D. degrees from Osaka University in 2005 and 2010, respectively. He is currently working as a postdoctoral researcher in the Laboratory for Image and Media Understanding, Kyushu University. His research interests include pattern recognition, gait motion recognition, robot localization and 3D map building, omnidirectional vision, and statistical robust estimation methods.

Yasushi Makihara received the B.S., M.S., and Ph.D. degrees in engineering from Osaka University in 2001, 2002, and 2005, respectively. He is currently an Associate Professor of the Institute of Scientific and Industrial Research, Osaka University. His research interests are computer vision, pattern recognition, and image processing including gait recognition, pedestrian detection, morphing, and temporal super resolution. He is a member of IPSJ, RJS, and JSME.

Hajime Nagahara received the B.E. and M.E. degrees in electrical and electronic engineering from Yamaguchi University in 1996 and 1998, respectively, and the Ph.D. degree in system engineering from Osaka University in 2001. He was a research associate of the Japan Society for the Promotion of Science (2001–2003). He was an assistant professor at the Graduate School of Engineering Science, Osaka University, Japan (2003–2010). He was a visiting associate professor at CREA University of Picardie Jules Verns, France, in 2005. He was a visiting researcher at Columbia University in 2007–2008. Since 2010, he has been an associate professor in faculty of information science and electrical engineering at Kyushu University. Computational photography, computer vision, and virtual reality are his research areas. He received an ACM VRST2003 Honorable Mention Award in 2003 and IPSJ Nagao Special Researcher Award in 2012.

Yasuhiro Mukaigawa received his M.E. and Ph.D. degrees from University of Tsukuba in 1994 and 1997, respectively. He became a research associate at Okayama University in 1997, an assistant professor at University of Tsukuba in 2003, an associate professor at Osaka University in 2004, and a professor at Nara Institute of Science and Technology (NAIST) in 2014. His current research interests include photometric analysis and computational photography. He is a member of IEEE.

Yasushi Yagi received the Ph.D. degree from Osaka University in 1991. He is the Director of the Institute of Scientific and Industrial Research, Osaka University, Ibaraki, Japan. In 1985, he joined the Product Development Laboratory, Mitsubishi Electric Corporation, where he worked on robotics and inspections. He became a Research Associate in 1990, a Lecturer in 1993, an Associate Professor in 1996, and a Professor in 2003 at Osaka University. International conferences for which he has served as Chair include: FG1998 (Financial Chair), OMINVIS2003 (Organizing chair), ROBIO2006 (Program co-chair), ACCV2007 (Program chair), PSVIT2009 (Financial chair), ICRA2009 (Technical Visit Chair), ACCV2009 (General chair), ACPR2011 (Program co-chair), and ACPR2013 (General chair). He has also served as the Editor of IEEE ICRA Conference Editorial Board (2007–2011). He is the Editorial member of IJCV and the Editor-in-Chief of IPSJ Transactions on Computer Vision & Applications. His research interests are computer vision, medical engineering, and robotics. Dr. Yagi was awarded theACMVRST2003 Honorable Mention Award, IEEE ROBIO2006 Finalist of T.J. Tan Best Paper in Robotics, IEEE ICRA2008 Finalist for Best Vision Paper, MIRU2008 Nagao Award, and the PSIVT2010 Best Paper Award. He is a fellow of IPSJ and a member of IEICE and RSJ.