21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Inertial-sensor-based Walking Action Recognition using Robust Step Detection and Inter-class Relationships

Ngo Thanh Trung, Yasushi Makihara, Hajime Nagahara,
Yasuhiro Mukaigawa, and Yasushi Yagi*

Abstract

This paper tackles a challenging problem of inertial sensor-based recognition for similar walking action classes. We solve two remaining problems of existing methods in the case of walking actions: action signal segmentation and recognition of similar action classes. First, to robustly segment the walking action under drastic changes such as speed, intensity, or style, we rely on the likelihood of heel strike that is computed employing a scale-space technique. Second, to improve the classification performance with similar action classes, we incorporate the inter-class relationship. In experiments, the proposed algorithms were positively validated with 97 subjects and five similar walking action classes, namely walking on flat ground, up/down stairs, and up/down a slope.

1 Introduction

Wearable and portable electronic devices are increasingly becoming useful to human life. They have rapidly become more and more sophisticated such that they interact or communicate with their users and understand the actions, needs, and health conditions of their users. With advances in micro-sensor and wireless communication technology, inertial sensors are now low-cost, lower-power, accurate, small, and effective. They are increasingly being embedded in such devices as smart phones. Therefore, many researchers have recently studied human assistance employing a wearable inertial sensor. Recognizing user actions through the inertial sensor is an important task for such assistance.

There have been a number of research papers on action recognition using wearable inertial sensors, which mainly differ in signal segmentation, feature extraction from a segment, and recognition technique for selected features. There are excellent reviews and comparisons of existing methods [1, 9].

For signal segmentation for feature extraction, a fixed-size sliding window has frequently been used [7, 8, 4, 12]. However, 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/user difference. The dynamic window [6] was proposed to solve the problem of the fixed-size window. This method relies on signal events detected according to a fixed threshold of the signal intensity to control the size and location of the window. Therefore, it faces a similar problem as for the fixed-size window because the signal intensity of an action can also vary. Moreover, in the field of gait-based user recognition, a number of algorithms that detect the walking step or cycle [10] are regarded also as dynamic window-based methods. However, they are not effective for a sequence of varying walking action. Therefore, existing methods have the problem of signal segmentation. Although dynamic time warping (DTW) can solve the problem of temporal variation, segmenting the signal for the action templates (or motifs) remains an unsolved problem [5].

There are various approaches for action recognition [8, 3, 12, 5]. However, 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. For such cases, the inter-class difference may be overwhelmed by the intra-class difference so that recognition is difficult.

In this paper, we tackle the above two problems in the case of recognizing walking action. *First*, the walking signal is segmented into steps employing a scale-space technique. The proposed step detection method can adaptively work with a large amount of variation even if the subject changes the walking speed or style. *Second*, 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 stable and distinguished patterns as in the case of walking action. We use these relationship patterns to recognize walking action.

*Ngo Thanh Trung, Yasushi Makihara, Yasuhiro Mukaigawa and Yasushi Yagi are with Osaka University, Japan.
Hajime Nagahara is with Kyushu University, Japan.

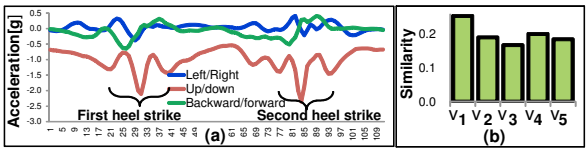


Figure 1. A real acceleration signal example of a walking cycle (a) and its feature vector (b).

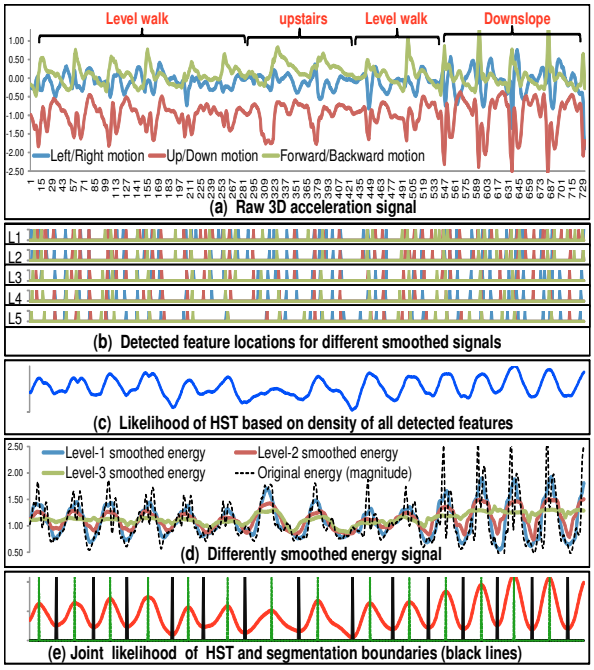


Figure 2. Example of the proposed step detection and segmentation algorithm. All graphs have the same temporal axis, and (a) and (b) use the same legends.

2 Proposed Method

A subject is assumed to walk on varying ground with a three dimensional (3D) accelerometer attached to his/her back waist. From a sequence of the captured signal, the segmentation of steps is performed automatically, and walking action samples are then made for two continuous steps. To recognize each test sample, we first compute a feature vector that describes the relationship to all of the walking action classes in a collection of action templates. This vector is then classified using a classifier such as the k-nearest neighbor (kNN) or support vector machine (SVM). In the following section, we briefly describe the ideas for the proposed algorithms.

2.1 Robust Step Detection

It is well known that a walking cycle consists of a stance phase and a swing phase [11]. For normal human walk, when a heel hits the ground at the start of the stance phase, the other foot remains on the ground. The impulse of the collision force is transmitted from the foot to the body center through the leg, which results in quick motion of the body center. Therefore, a 3D accelerometer attached at the back waist can capture a strong signal at the moment of the *heel strike* (HST). Within a walking cycle, we can observe strong signal vibration at two such moments for the two legs, as illustrated in Fig. 1(a). We use this phenomenon to extract the signal segment of a step and a walking cycle relying on the computation of the likelihood of HST.

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

- $Obsv1$: The density of local feature points (e.g., peaks and valleys) in all channels is relatively high,
- $Obsv2$: Energy of the acceleration signal is relatively high.

Based on $Obsv1$, we use locations of local peaks and valleys for each channel of the signal as the signal features. To robustly compute the feature density against temporal variation and noise, we employ a scale-space technique. First, the 3D signal sequence, illustrated in Fig. 2(a), is smoothed by several Gaussian filters with different smoothness scales. We then detect all the signal features for each channel and each smoothness scale, as illustrated in Fig. 2(b). Finally, from all the detection results, the probability density function of features p_t^f at time t is computed by kernel density estimation, Fig. 2(c).

Based on $Obsv2$, we regard the energy of the acceleration signal as another likelihood of HST. Energy e_t at time t is computed as the magnitude of the 3D signal $\boldsymbol{s}_t$: $e_t = ||\boldsymbol{s}_t||$. For robustness against temporal variation and noise, we compute several smoothed signal energies $\hat{e}_{w_l,t}$ with smoothing parameter w_l of smoothness level l, Fig. 2(d). The likelihood of HST based on signal energy is: $p_t^e = \rho \prod_l \hat{e}_{w_l,t}$, where ρ is a scale factor.

Considering both $Obsv1$ and $Obsv2$, the likelihood of HST p_t is computed as the product of two likelihoods p_t^f and p_t^e:

$$p_t = p_t^f p_t^e. \quad (1)$$

Because the HST should contain meaningful information for classifying actions, it would be better to segment the signal into steps so that the HSTs are located at the center of the segmented steps rather than at the segmentation boundaries. The local peaks of p_t, illustrated by dashed green lines in Fig. 2(e), are considered as approximations for the HST locations. A minimum local valley between two adjacent local peaks is used as

the segmentation location. Action steps are then segmented by all these local valleys, as illustrated by black lines in Fig. 2(e).

Action samples for recognition are constructed for two consecutive steps.

2.2 Recognition using Inter-class Relationship

A set of action templates $\mathbb{G}$ is constructed using action samples generated by training sequences for various subjects: $\mathbb{G} = \{\boldsymbol{G}_i | i = 1 \dots n\}$, where $\boldsymbol{G}_i$ is a collection of action class i and n is the number of classes. In the recognition of a test sample $\boldsymbol{p}$, the proposed recognition method uses the inter-class relationship patterns to improve the recognition performance. The method involves two steps: *representation* and *recognition*.

In the first step, a feature vector $\boldsymbol{v_p} = [v_1, \dots, v_n]^T$, describing all the intra-class and inter-class relationships, is computed for a test sample $\boldsymbol{p}$, where v_i is the similarity that $\boldsymbol{p}$ belongs to class i and $\sum_{i=1}^{n} (v_i)^2 = 1$. Obviously, we can obtain the classification result at this point by simply selecting the class with highest similarity. An example is shown in Fig. 1(b) for $\boldsymbol{v_p}$, where the test sample is classified as action 1 since v_1 is the largest. However, we continue to use $\boldsymbol{v_p}$ for further classification to improve the result in the case of walking actions. In our algorithm, action samples are simply raw 3D signals, different in size, and the distance $d(\boldsymbol{p}, \boldsymbol{g})$ between $\boldsymbol{p}$ and each $\boldsymbol{g} \in \boldsymbol{G}_i$ is computed:

$$d(\boldsymbol{p}, \boldsymbol{g}) = DTW(\boldsymbol{p}, \boldsymbol{g}), \tag{2}$$

where $DTW(.)$ is a DTW function that returns the distance between two sequences. The distance $D(\boldsymbol{p}, \boldsymbol{G}_i)$ between $\boldsymbol{p}$ and template action class $\boldsymbol{G}_i$ is computed as the average distance:

$$D(\boldsymbol{p}, \boldsymbol{G}_i) = \frac{1}{k} \sum_{\boldsymbol{g} \in kNN(\boldsymbol{p}, \boldsymbol{G}_i)} d(\boldsymbol{p}, \boldsymbol{g}), \tag{3}$$

where $kNN(\boldsymbol{p}, \boldsymbol{G}_i)$ is a function that returns a set of k nearest neighbors of $\boldsymbol{p}$ in $\boldsymbol{G}_i$.

The feature vector $\boldsymbol{v_p} = [v_1, \dots, v_n]^T$ is then converted from $[D(\boldsymbol{p}, \boldsymbol{G}_1), \dots, D(\boldsymbol{p}, \boldsymbol{G}_n)]^T$. Each template sample $\boldsymbol{g} \in \boldsymbol{G}_i$ is also used as an input in the first step to compute the output $\boldsymbol{v_g}$ in a leave-one-out manner. In other words, $\boldsymbol{v_g}$ is computed when it is excluded from $\mathbb{G}$.

Once a training data set of n-dimensional vectors $\boldsymbol{v_g}$ is prepared, a classifier such as SVM or kNN is constructed, and the action associated with feature vector $\boldsymbol{v_p}$ of input test sample $\boldsymbol{p}$ is then classified.

A flowchart of the proposed two-step recognition algorithm is presented in the Fig. 3.

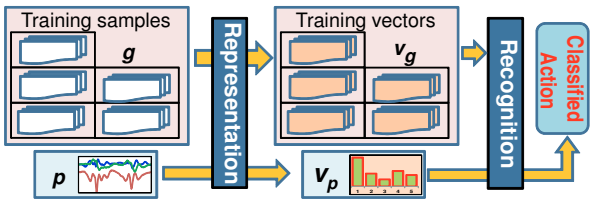
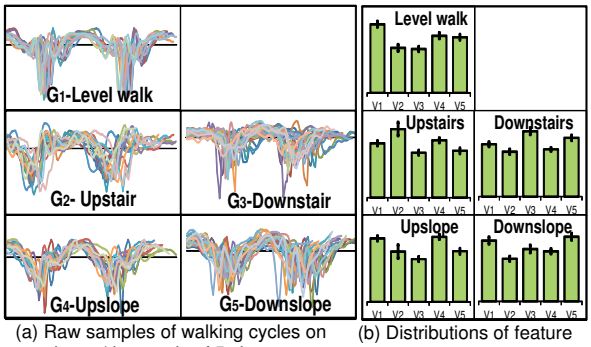


Figure 3. The proposed recognition algorithm consists of two steps: *representation* and *recognition*.



(a) Raw samples of walking cycles on the up/down axis of 5 classes

(b) Distributions of feature vectors of 5 classes

Figure 4. Walking action samples of training data set (a) and distributions of their feature vectors (b).

3 Experiments

In our experiments, an accelerometer was attached to the back waist of a subject and captured data at a sampling period of 10 ms. Each subject was asked to walk across straight flat ground, up stairs, down stairs, up a slope, and down a slope freely in the same environment. Ground truth action labels were assigned manually by synchronizing with simultaneously captured video. Meanwhile, step detection was executed automatically and walking cycles were segmented for action samples. We collected data from 97 subjects aged 15 to 70 years. The subjects were separated randomly into 53 training and 44 test subjects.

An example of data segmentation for action samples is shown in Fig. 4(a) for five action classes of the whole training set. In Fig. 4(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, respectively. We see that the walking-cycle segmentation worked well and the inter-class relationships have clear and relatively distinguished patterns for each action class, which strongly encourages the use of the proposed recognition algorithm.

We compared the proposed method with a bag-of-features method, which uses fixed-size window for sig-

nal segmentation [12], denoted BOF2012. 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 to compare with the proposed method, 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. We also evaluated the result at the early stage of the proposed recognition algorithm before using the inter-class relationship by selecting the highest similarity for each test sample, which is denoted NO_CR. In the case of using the proposed feature vector, we compared the results of two classifiers, kNN and SVM [2], which are respectively denoted PROPOSED_KNN and PROPOSED_SVM. For the SVM, the option of multiple binary classifiers with a linear kernel was selected. The accuracies for all the action classes and their average are shown in Tab. 1 for each method.

From the results in Tab. 1, we see that walking up/down a slope is the most difficult action to be recognized. The reason is that walking up/down a slope is easily confused with walking on flat ground or up/down stairs, meanwhile walking on flat ground and walking up/down stairs are quite distinguished. Compared with BOF2012, the proposed method with the proposed step detection algorithm is overall effective even without using inter-class relationship information. The reason is that BOF2012 uses a fixed-size window for signal segmentation, it cannot segment an action accurately crossing different speeds and subjects, while the proposed step detection can. Moreover, the proposed method does not absolutely require user-defined parameters to work. From the results of PROPOSED_SVM and PROPOSED_KNN compared with that of NO_CR, we see that inter-class relationship information is useful and improves the recognition performance.

4 Conclusion

We proposed a recognition method for similar walking actions using an accelerometer. We proposed a robust step detection method to segment a signal into action samples. The method works well even if the action drastically varies in speed or intensity. We also proposed a recognition method using a feature vector composed of similarities to all action classes to improve the performance compared with the case of using a single similarity for a target action class. Experiments for five walking action classes (walking on flat ground, up stairs, down stairs, up a slope, and down a slope) positively validated the proposed method.

Table 1. Accuracy comparison(%)

Method	Level walk	Up stairs	Down stairs	Up slope	Down slope	**Average**
NO_CR	96.9	93.2	85.4	70.8	80.6	85.4
PROPOSED_KNN	94.0	***95.5***	***95.1***	71.7	85.8	88.4
PROPOSED_SVM	***97.0***	92.9	***95.1***	***78.3***	***88.8***	***90.4***
BOF2012	90.3	95.2	90.2	47.5	69.0	78.5

Acknowledgment

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

References

[1] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. J. M. Havinga. Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey. In *Intl. Conf. on Architecture of Computing Systems, Workshop Proc.*, 2010.

[2] C. C. Chang and C. J. Lin. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Tech.*, 2011.

[3] H. Ghasemzadeh, V. Loseu, and R. Jafari. Structural action recognition in body sensor networks: Distributed classification based on string matching. *Infor. Tech. in Biomedicine, IEEE Trans. on*, 14(2):425 –435, 2010.

[4] C. W. Han, S. J. Kang, and N. S. Kim. Implementation of hmm-based human activity recognition using single triaxial accelerometer. *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, E93.A(7):1379–1383, 2010.

[5] B. Hartmann and N. Link. Gesture recognition with inertial sensors and optimized DTW prototypes. In *Systems Man and Cybernetics, IEEE Intl. Conf. on*, 2010.

[6] J. O. Laguna, A. G. Olaya, and D. Borrajo. A dynamic sliding window approach for activity recognition. In *Proc. of the 19th intl. conf. on User modeling, adaption, and personalization*, pages 219–230, 2011.

[7] N. Lovell, N. Wang, E. Ambikairajah, and B. Celler. Accelerometry based classification of walking patterns using time-frequency analysis. In *Eng. in Medicine and Biology Society, Intl. Conf. of the IEEE*, 2007.

[8] A. Mannini and A. Sabatini. On-line classification of human activity and estimation of walk-run speed from acceleration data using support vector machines. In *Engineering in Medicine and Biology Society,EMBC, Annual Intl. Conf. of the IEEE*, pages 3302 –3305, 2011.

[9] S. Preece, J. Goulermas, L. Kenney, and D. Howard. A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *Biomedical Engineering, IEEE Trans. on*, 56(3):871 –879, march 2009.

[10] N. Trung, Y. Makihara, H. Nagahara, R. Sagawa, Y. Mukaigawa, and Y. Yagi. Phase registration in a gallery improving gait authentication. In *Proc. of the Intl. Joint Conf. on Biometrics*, 2011.

[11] C. L. Vaughan, B. L. Davis, and J. C. O. Connor. *Dynamics of Human Gait*, chapter 2. Kiboho Publishers, 2nd edition, 1999.

[12] M. Zhang and A. A. Sawchuk. Motion primitive-based human activity recognition using a bag-of-features approach. In *Proc. of the 2nd ACM SIGHIT Intl. Health Informatics Symposium*, 2012.