

Automatic Labeling Framework for Wearable Sensor-based Human Activity Recognition

Guanhao Liang,¹ Qingsheng Luo,^{1*} and Yan Jia²

¹School of Mechatronical Engineering, Beijing Institute of Technology,
No. 5 Zhongguancun South Street, Haidian District, Beijing 100081, China

²School of Mechanical Engineering, Beijing Institute of Technology,
No. 5 Zhongguancun South Street, Haidian District, Beijing 100081, China

(Received April 24, 2018; accepted June 25, 2018)

Keywords: human activity recognition, wearable sensors, labeling effort reduction, endpoint detection, automatic labeling

Labeled datasets are one of the key factors for obtaining a good and robust classifier using supervised learning methods. However, labeling raw data is a tedious and labor-intensive process, which is usually done manually. Many efforts were proposed to utilize a small amount of labeled data to train a classifier that is sufficiently robust to label more data for training or make a prediction on unlabeled data. Unlike previous studies, we proposed an automatic labeling framework without labeling a small amount of data in advance, to directly annotate unlabeled time series data regarding body-worn sensor-based human activity recognition (HAR) in laboratory settings. The framework automatically labels collected time series activity data by transforming the original data into its corresponding absolute wavelet energy entropy and detects activity endpoints based on constraints and information extracted from a predefined human activity sequence. The performance of the proposed framework was evaluated on the collected dataset and the UCI HAR Dataset. In both cases, the average precision and recall scores are above 81.9% and the average F-measure scores are above 88.9%. Results showed that the proposed framework can be adopted as a rapid and reliable way of generating labeled datasets from unlabeled data.

1. Introduction

Since the 1990s, researchers have begun to use wearable sensors for human activity recognition (HAR).⁽¹⁾ From the perspective of sensor types, research studies on HAR mainly include vision sensors (such as cameras), ambient sensors in smart home scenes, and wearable sensors [such as accelerometers, gyroscopes, and inertial measurement units (IMUs)], which mainly apply supervised learning methods to learn different human activity patterns from collected human motion data. Therefore, acquiring a proper set of labeled data is the basis for training HAR models. Annotation techniques can be generally classified as offline and online methods. Specifically, offline methods include self-recall,⁽²⁾ indirect observation,

*Corresponding author: e-mail: luoqsh@bit.edu.cn
<https://dx.doi.org/10.18494/SAM.2018.1962>

and video and audio recordings,^(3,4) and online methods include direct observations,⁽⁵⁾ time diary, and experience sampling.^(6,7) For studies acquiring data in a laboratory scenario, direct observation and video recordings are usually taken as the annotation methods, which might also be called on-site annotation and *post hoc* annotation, respectively.⁽⁸⁾ For the former method, an annotator records the timestamp range of human activity currently performed by a subject. For the latter method, an annotator needs to compare between the video footage of the whole data acquisition process and the acquired time series to complete raw data annotation.^(9,10) For long-term HAR studies, especially focusing on activity monitoring, online methods are more appealing for realistic application whereas offline methods almost make it impossible to obtain ground truth labels, which are always labor-intensive and usually unacceptable due to privacy concerns.⁽¹¹⁾ There is a tradeoff between the accuracy of an annotation method and the time and effort required for annotation. Offline methods can provide more accurate annotations than online methods but demand massive effort, especially when the set of data is large. Although online methods are less time-consuming, inaccurate annotations and more ambiguity may be introduced to the labeled dataset.

In this work, we target supervised learning-based applications that mainly acquire data in laboratory settings. In such controlled settings, annotations can often be obtained by video recordings or direct observations. For annotators, video recordings are generally easier to interpret than time series data, and annotation efforts will increase when the numbers of activities and subjects become larger and the data acquisition time becomes longer. To obtain detailed annotations with acceptable accuracy and reduce the annotation efforts in *post hoc* labeling settings, we proposed a novel automatic labeling framework (ALF) towards multivariate time series data acquired from multiple wearable sensors. Unlike previous methods based on machine learning, we tackle this problem from a speech processing perspective. The proposed framework consists of two main steps: (1) inserting a rest position between two human activities during data acquisition to create a human activity sequence (HAS), and (2) extracting a wavelet energy entropy (WEE) feature from HAS and detecting endpoints of each human activity. The proposed ALF can accelerate the process and reduce the cost of data annotation.

The rest of the paper is organized as follows. In Sect. 2, we provide a brief overview of related work on reducing labeling efforts. In Sect. 3, the proposed ALF is demonstrated in detail. Experimental results of the proposed ALF are discussed in detail in Sect. 4. Finally, conclusions are presented in Sect. 5.

2. Related Work

Supervised algorithms with high recognition performance usually require significant amounts of labeled training data. In previous HAR studies based on wearable sensors, especially for those using supervised algorithms, manual methods, e.g., video recordings and direct observations, were widely used to obtain annotations. Plotnik *et al.* developed a wearable assistant for Parkinson's disease patients with freezing of gait (FOG) symptom.⁽¹²⁾ In the data acquisition phase, two annotators were assigned to conduct on-site annotation. One was using a digital video camera to record subjects' activities including standing, walking, turning, and

freezing, while the other was assigning the corresponding labels in real time to the acceleration data transmitted from a wearable assistant on a laptop. A physiotherapist then determined the endpoints of FOG events in the acquired data according to a *post hoc* analysis on the video recordings. Similarly, Anguita *et al.* used a smartphone (Samsung Galaxy S II) to collect human motion data and perceived different types of human activities on the basis of ambient information.⁽¹³⁾ Acceleration and angular velocities of typical daily activities, including standing, sitting, lying, walking, going downstairs, and going upstairs, were collected from 30 subjects aged 19 to 48 years. Each subject was requested to perform two rounds of each set of human activities and to rest for 5 s between each round. Acquired data were manually labeled afterwards according to the video footage of subjects' activities. Banos *et al.* attached two IMUs to each subject's right wrist and left ankle and another sensor to the chest, which provides two-lead ECG measurements.⁽¹⁴⁾ Thus, acceleration, angular velocity, geomagnetic information, and ECG of 12 different outdoor human activities were collected among 10 volunteers. The entire data acquisition process was recorded by a video camera and then manually annotated. Furthermore, there are many other public HAR datasets based on wearable sensors or portable devices available online in which acceleration, angular velocity, and geomagnetic signals are mostly collected, and some vital signals are also acquired for future work purposes.^(15–18) Detailed annotation with high accuracy can be obtained by manual annotation methods with intensive labeling efforts.

In long-term human activity monitoring settings, supervised algorithms face the challenge of obtaining a labeled dataset. To cope with the problem of insufficient labeled data, some researchers have moved from fully supervised settings to weakly supervised ones so as to reduce annotation efforts. By incorporating with experience sampling, multi-instance learning (MIL) obtains knowledge from a significantly weak labeled dataset in which labels are associated with sets (bags) of instances, instead of training instances. In this way, sensor data can be labeled on a very coarse level, which significantly lowers the annotation burden. A bag is labeled positive if and only if at least one positive instance, i.e., the activity we are interested in, exists in the bag, and negative if all instances in the bag are negative. The first work of MIL on time series data for HAR was Ref. 11 and it adapted modifications of Support Vector Machine (mi-SVM) to three different bag-labeling scenarios (i.e., single-labeled bags, multilabeled bags, and majority-voting bags). The extensive study and comparative evaluation proved the capability of significantly decreasing the annotation efforts in the proposed MIL-based methods. Guan *et al.* proposed a novel MIL model based on the work described in Ref. 11 for offline activity recognition from multivariate time series data, and it is a generative graphical model based on an Auto-Regressive Hidden Markov Model (HMM), which can predict both bag and instance labels.⁽¹⁹⁾

Unsupervised learning techniques are usually used to discover the underlying structures in activity data without the necessity of providing labels. Wyatt *et al.* viewed activity data as a stream of natural language terms, i.e., sequences of object use, and generic models mined by daily activities from the web, which served as common sense in HAR.⁽²⁰⁾ Bottcher *et al.* proposed an unsupervised framework of adopting clustering algorithms to detect transitions between steps of manual work that follows a (semi-) fixed procedure. Although the order and/or

number of steps in the process may be given in advance, the framework removes the necessity for labeled data.⁽²¹⁾ By using unsupervised techniques, the effort of labeling activity data is unnecessary.

Transfer learning is defined as the ability to extend what has been learned in one context to new contexts and relies on the assumption that some underlying relationship between the source and the target areas exists and allows for the successful transfer of knowledge from the source to the target.^(22,23) In a smart home setting, Kasteren *et al.* proposed three different function groups to project extracted features to a common space and then used both unlabeled data from house A and labeled data from house B to learn the parameters of a semisupervised HMM for activity recognition in house A.⁽²⁴⁾ In another work, Kasteren *et al.* transferred the knowledge obtained by using existing labeled data from various homes to an HMM model applied in a new home.⁽²⁵⁾ Chen *et al.*, however, proposed a transfer learning framework based on principal component analysis (PCA) transformation, Gale–Shapley similarity measurement, and Jensen–Shannon divergence (JSD) feature mapping.⁽²⁶⁾

Semisupervised learning makes use of only a small amount of labeled training data and a substantial amount of unlabeled training data.⁽²⁷⁾ For example, self-training, cotraining,⁽²⁸⁾ and En-Co-Training are some typical semisupervised techniques whereas a special case of semisupervised learning,⁽²⁹⁾ i.e., active learning, mainly focuses on labeling the most profitable instances, but human intervention is necessary to some extent for a small amount of labeled data. For example, Zhao *et al.* proposed a robust active learning model using crowdsourced annotations for activity recognition.⁽³⁰⁾

The aforementioned methods addressed the issue of data annotation by mainly using two kinds of annotation techniques, which adopt either intensive labeling efforts or learning methods. As the research moves from a laboratory setting to a real-world setting, detailed labeled data are more difficult to obtain. As a result, using a small amount of labeled data and prior knowledge of the target activities to train a learning model that can classify activity of daily life (ADL) with acceptable accuracy is the core idea of MIL, unsupervised learning, transfer learning, semisupervised learning, and active learning. However, manual labeling of the initial data is required to some extent by these methods. Thus, we propose a framework to achieve automatic annotation of time series activity data in a laboratory setting, which is based on the prior knowledge of the acquired data and the WEE feature to automatically detect endpoints of human activities.

3. Proposed ALF

The fact that it is difficult to interpret time series data generated by wearable sensors, such as IMUs, makes it necessary to refer to video recordings of the data acquisition process when the manual annotation is conducted so as to guarantee labeling accuracy. Video recording methods are still widely used in data annotation in a laboratory setting when supervised learning methods are adopted to train an activity classifier because it can be seen as a kind of prior knowledge of the acquired data, which can help annotators interpret time series data during labeling. However, when the number of subjects taking part in the data acquisition increases or

the duration of data acquisition per subject becomes longer, labeling efforts would be intensive and time-consuming. Automatic methods to address this issue should take into account the prior knowledge of acquired data. In this work, we consider the information contained in the acquired sequence of human activities as a type of prior knowledge of time series data, e.g., the order of activities, the longest/shortest duration of acquired activities, the longest/shortest duration of rest posture (RP), and the possible lowest acceleration during an activity.

Therefore, we proposed an ALF including HAS, feature extraction, and automatic labeling as shown in Fig. 1. On one hand, HAS is predesigned as a data acquisition scheme for acquiring time series activity data. On the other hand, HAS provides prior knowledge, i.e., constraints and the activity sequence for automatic labeling. Details of each step regarding this framework are demonstrated in the following sections.

3.1 HAS

HAS is a set of time-aligned human activities that are predesigned before conducting the data acquisition process. In the real world, the boundary between different human activities is not distinct, which makes it challenging to segment two different but consecutive human activities.⁽¹⁾ For instance, Refs. 12 and 18 collected consecutive human activities with no apparent pause state between every two human activities. There are mainly two different ways of labeling a time series data. One is to label every single activity that is performed by each subject while the other is to label specific activities of interest. Figure 2(a) is subject 1's acceleration and corresponding labels in Ref. 13. Each human activity (i.e., walking, ascending stairs, descending stairs, sitting, standing, and lying down) is labeled as a decimal number from one to six, respectively. Figure 2(b) shows the acceleration of subject 2's activity during data acquisition in Ref. 14. Only activities of interest (i.e., standing and ascending stairs) are labeled. When special locations are required to conduct data acquisition for different kinds of activities, irrelevant activities were created and labeled as number zero between two different activities of interest, e.g., the subject needs to move from the laboratory to a building with stairs so as

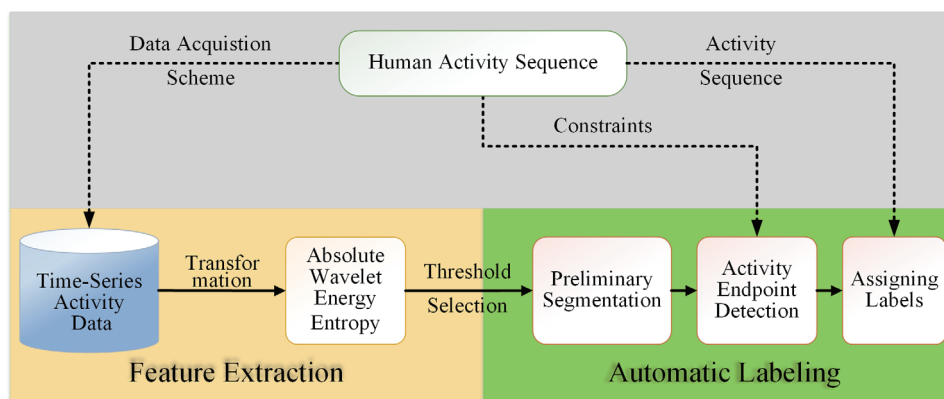


Fig. 1. (Color online) Proposed ALF.

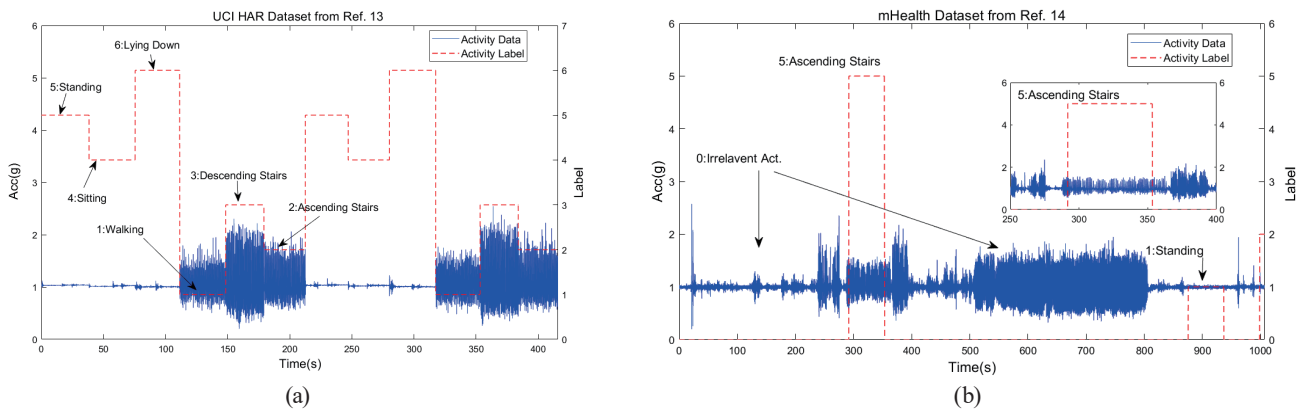


Fig. 2. (Color online) (a) UCI HAR Dataset from Ref. 13 and (b) segment of mHealth Dataset from Ref. 14.

to ascend the stairs. Lee and Xu⁽³¹⁾ and Amft *et al.*⁽³²⁾ introduced a predefined rest position between two different hand gestures in order to create distinct boundaries for different types of hand gestures. In this work, we inserted RP between each target activity performed by each subject during data acquisition in order to make the boundaries of human activities clearer for labeling. An RP is defined as the stationary status of a subject during data acquisition. For example, standing still, sitting, and lying down are three typical RPs a subject can simulate.

Therefore, we designed the HAS as shown in Fig. 3. It groups the activities of normal people in daily life into three different types, namely, stationary (STACT), quasi-periodic (QPACT), and sporadic (SPACT) activities. STACTs indicate that the subject stays in a static posture, i.e., standing still (STND), sitting (SIT), lying down (LD), and RP. QPACTs show a pattern of recurrence with similar movements during a subject's activities such as walking (WLK), ascending (AS) and descending stairs (DS), jumping up and down (JUD), and squatting slowly (SS) and quickly (SQ). SPACTs indicate activities that occur sparsely or accidentally in our daily life with uncertainty such as falling, which is rare but of high risk. Particularly, we simulated two kinds of falls, i.e., falls with spontaneous protection (SP) and those without SP. A simulated fall with SP means that the subject spontaneously triggers self-protective actions, e.g., bending knees or stretching arms straight to mitigate the impact between the subject's body and the foam, to protect self from an impending and inevitable fall. In contrast, a simulated fall with no spontaneous protection (NSP) simulates a state when one is unconscious (e.g., faint or falling asleep) or being in a pathological state (e.g., having a syncope or hemiplegia) and failing to spontaneously trigger self-protective actions during a fall. Each type of fall is further categorized into four types of specific falls on the basis of the direction of falling, i.e., fall forward (FF), fall backwards (FB), fall to the left (FL), and fall to the right (FR). In total, time series data of 17 target activities are collected and denoted as decimal numbers from 1 to 17.

To date, some hardware and sensors have been adopted to acquire human motion data. These devices can be categorized into two types, namely, commercial products (Samsung Galaxy S2, HTC Magic, SHIMMER, and Xsens) and research prototypes (Wocket, 3dNX, GENE, and e-AR).^(33–40) Commercial products such as a smartphone or a smartwatch may

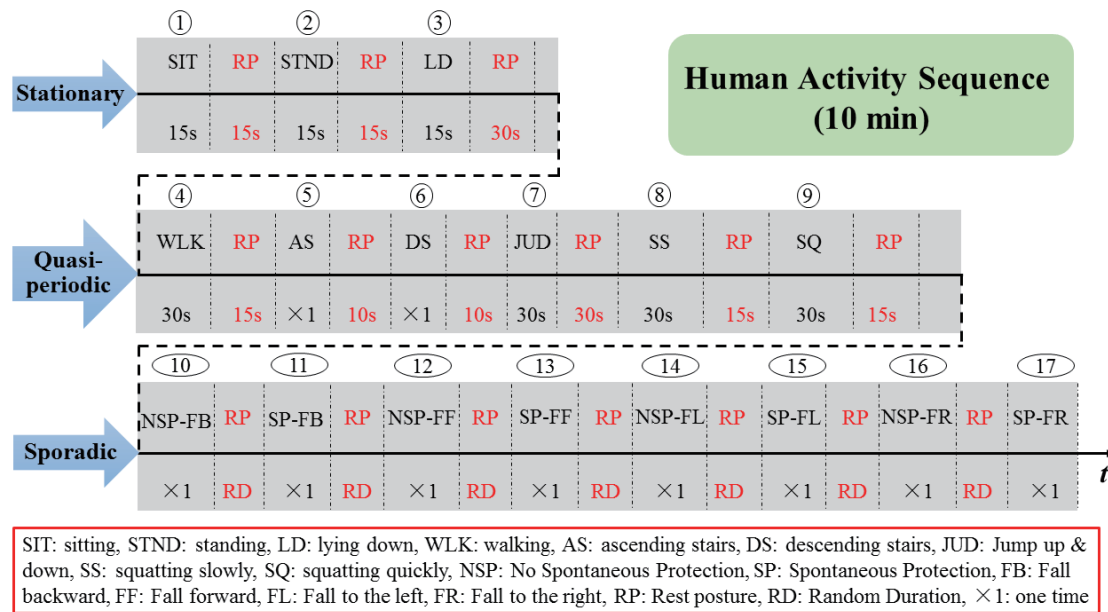


Fig. 3. (Color online) Proposed HAS with a duration of 10 min.

be uncomfortable to be attached to places other than wrists of a wearer while products such as SHIMMER and Xsens are expensive. Moreover, research prototypes developed by other researchers may be unavailable for sale or to obtain. To seek a low-cost miniature hardware that acquires motion data of human activities, we developed IMU modules, each of which is based on a Microcontroller Unit (MCU), STM32F103R, and a six-axis MEMS motion tracking device, MPU6050, which consists of a tri-axial accelerometer and a tri-axial gyroscope, as shown in Fig. 4(a). Nine IMUs were attached to each subject on nine different locations using hook-and-loop fasteners as shown in Fig. 4(b), i.e., left upper arm (LUA), right upper arm (RUA), left wrist (LWR), right wrist (RWR), front waist (FW), left thigh (LT), right thigh (RT), left ankle (LAK), and right ankle (RAK), as shown in Fig. 5(a). Data acquisition was conducted in a laboratory setting as shown in Fig. 5(b). Six subjects (five males and one female) were chosen from students aged between 26 and 28. Each subject was asked to follow the experimenter's instructions to start or terminate a specific activity and to perform each of them thoroughly and completely according to their understanding of each activity. Also, slow and steady moves were mostly recommended to subjects during transition activities. Particularly, all simulated falls are self-initiated by each subject and all subjects are protected by a foam after a simulated fall. No further instruction on how to perform each activity is provided.

During the process of data acquisition, each subject was asked to start performing activity 1 at location A and to end up with activity 17 at location E sequentially according to HAS. Moreover, each subject shifted the location as target activities are being performed as shown in Fig. 5, i.e., to start with sitting at A → standing at B → lying down at E → walking steadily (B ↔ C) → ascending stairs (C → D) → descending stairs (D → C) → jumping up and down at C →

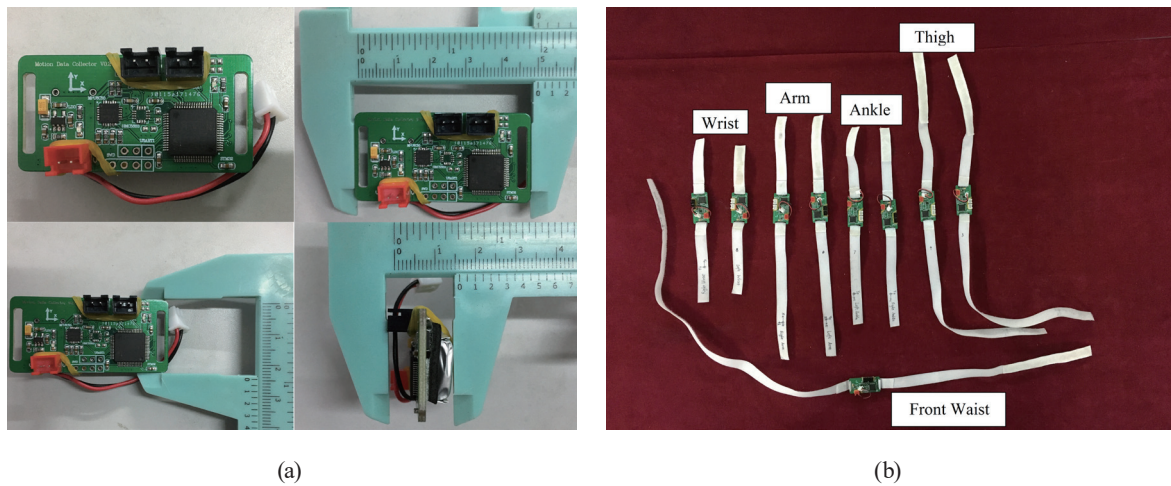


Fig. 4. (Color online) (a) An IMU module and (b) IMU modules with hook-and-loop fasteners.

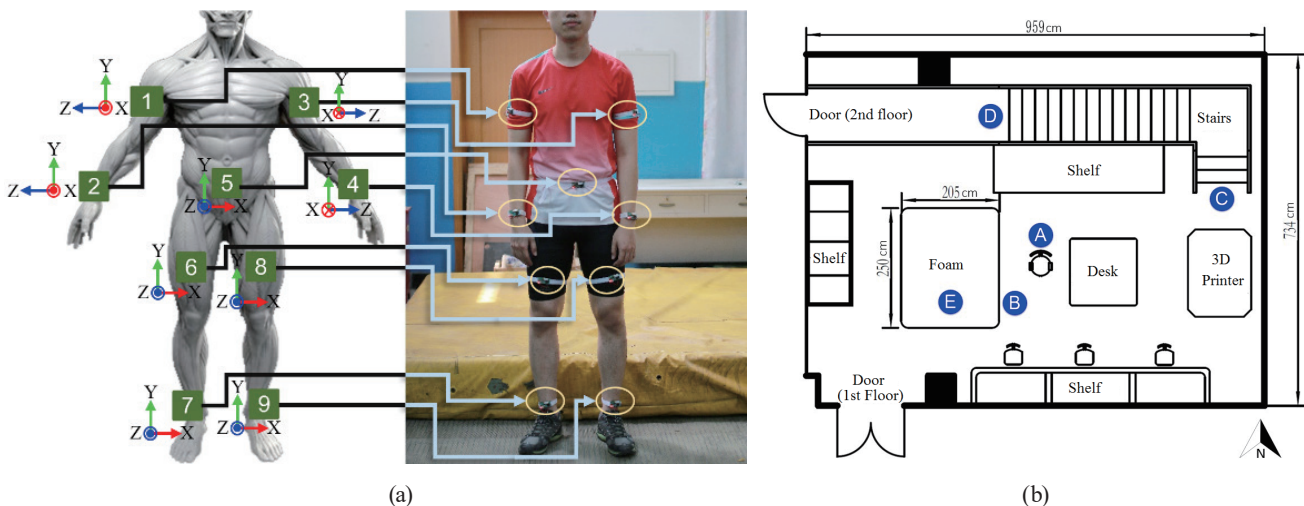


Fig. 5. (Color online) (a) Layout and coordinate systems of IMUs attached to a subject and (b) layout of the laboratory setting.

walking slowly to B and facing the 3D printer → squatting slowly at B → squatting quickly at B → NSP falling backwards (B → E) → SP falling backwards (B → E) → NSP falling forwards (B → E) → SP falling forwards (B → E) → NSP falling to the left (B → E) → SP falling to the left (B → E) → NSP falling to the right (B → E) → SP falling to the right (B → E). Each IMU module was designed to collect a stream of data with a duration of 10 min in a constant sampling rate of 100 Hz. As a result, a stream of activity data lasting 600 s was collected for each subject, and the durations of STACT, QPACT, and SPACT are approximately 105, 235 and 260 s, respectively. During data acquisition, some transition movements are inevitably created owing to the necessity of moving from one location to another so as to complete the whole process. Therefore, RP and transition movements should both be labeled as number zero in the following automatic labeling process. Figures 6(a) and 6(b) show two sets of FW acceleration

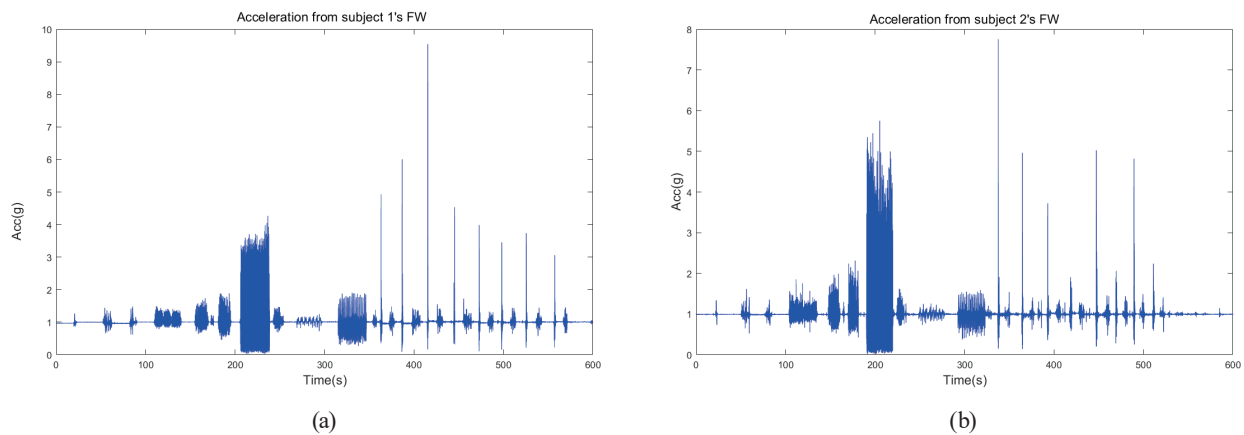


Fig. 6. (Color online) (a) Unlabeled subject 1's FW resultant acceleration and (b) unlabeled subject 2's FW resultant acceleration.

obtained from two subjects, the target activities of which have clearer boundaries when compared with Fig. 2. Note that the presented HAS is merely a scheme that cannot be seen as a direct source of providing ground-truth labels of the collected data owing to the fact that either a timekeeper or a subject cannot time or perform as promptly as the target activities planned in the HAS. That is, lead-lag timing or performing a target activity is inevitable and certainly has a lead-lag impact on the schedule of each activity to follow in the collected data.

3.2 Feature extraction

Various studies have been conducted on the endpoint detection regarding speech signals.^(41–43) The main task of endpoint detection of speech signals is to detect the start and end points of a speech signal. The collected acceleration data in Figs. 6(a) and 6(b) are similar in morphology to a speech signal. That is, the stationary part of acceleration can be treated as the mute part in a speech signal while the oscillating part of acceleration can be treated as the part containing a speech. In this work, we consider the processing of collected data in a speech processing perspective.

Many algorithms have been proposed to tackle the endpoint detection of speech signals, which mostly aim to extract different features from the original signal, e.g., spectral entropy, cepstrum distance, and dual thresholds. Among various endpoint detection techniques, energy-based methods are the most widely applied solutions to this problem.⁽⁴¹⁾ These algorithms are mostly based on the short-time Fourier transform (STFT). However, STFT has a fixed resolution, which might lead to poor time/frequency resolution in the analysis of nonstationary signals, e.g., a speech signal and a time series of acceleration collected here are both nonstationary signals. Moreover, the energies of different activities vary from each other. To locate the boundaries of activity data with better accuracy, we adopted multilevel one-dimensional (1-D) wavelet decomposition to the original activity data and extracted the corresponding detail coefficient of each level. In this way, the wavelet energy distribution along

different levels of decomposition can be calculated and described by the WEE feature.

Multilevel 1-D wavelet decomposition is based on wavelet transform. Suppose $\forall f \in L^2(\mathbb{R})$, its continuous wavelet transform (CWT) is defined as Eq. (1).

$$CWT_f(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(x) \overline{\psi\left(\frac{x-b}{a}\right)} dx = \langle f, \psi_{a,b} \rangle, \quad (1)$$

where the CWT of a given function $f(x)$ is equal to the inner product between $f(x)$ and the wavelet function $\psi_{a,b}$ in which $\psi_{a,b}$ is equal to $\psi\left(\frac{x-b}{a}\right)/\sqrt{|a|}$ ($a, b \in \mathbb{R}$). In practice, the discrete wavelet transform (DWT) is mostly adopted and its parameters usually take dyadic values, i.e., $a = 1/2^j$, $b = k/2^j$ ($j, k \in \mathbb{Z}$). Thus, the DWT of $f(x)$ is denoted as Eq. (2), which is a dyadic orthogonal wavelet transform to create wavelet basis functions for multiresolution analysis (MRA).

$$DWT_f\left(\frac{1}{2^j}, \frac{k}{2^j}\right) = \langle f, \psi_{j,k} \rangle \quad (2)$$

To apply MRA to $f(x)$ is to perform multilevel wavelet decomposition on $f(x)$ and to reconstruct $f(x)$ as Eq. (3).

$$f(x) = \sum_{k \in \mathbb{Z}} c_{j,k} \phi_{j,k}(x) + \sum_{i=1}^J d_{i,k} \psi_{i,k}(x) = \hat{f}(x) + \hat{g}(x), \quad (3)$$

where $\phi_{j,k}(x)$ are scale functions to compose the approximation item $\hat{f}(x)$, and $\psi_{i,k}(x)$ are wavelet functions to compose the detail item $\hat{g}(x)$. $c_{j,k} = \langle f, \phi_{j,k} \rangle$ are the approximation coefficients, $d_{i,k} = \langle f, \psi_{i,k} \rangle$ are the detail coefficients, and J is the number of levels of decomposition. Detail coefficients capture high-frequency components of the original signal. In this work, segments with oscillating acceleration contain high-frequency components while those with steady acceleration contain low-frequency components. Thus, the magnitude of detail coefficients indicates the energy of the signal segment. Thus, the wavelet energy of a J -level 1-D wavelet decomposition is defined as Eq. (4).

$$Et = EA_J + \sum_{i=1}^J ED_i = |c_{J,k}|^2 + \sum_{i=1}^J |d_{i,k}|^2, \quad (4)$$

where Et is the total wavelet energy, i is the i -th level wavelet decomposition ($i = 1, 2, \dots, J$), ED_i denotes the detail energy of the i -th level wavelet decomposition, and EA_J denotes the approximation energy of the J -th level wavelet decomposition. In combination with Shannon entropy, WEE indicates the distribution of human activities along all level wavelet decompositions and can be defined as Eq. (5).⁽⁴⁴⁾

$$Entropy_WE = -\sum_{i=1}^J p_i \log_2 p_i, \quad (5)$$

where p_i is the probability density of the i -th level wavelet energy, $p_i = ED_i/Et$ ($i = 1, 2, L, J$), and $Entropy_WE$ denotes WEE.

Assume the collected time series activity data as $f(x)$, which is a vector of dimension $M \times 1$. The sliding window technique is adopted to extract the WEE feature from $f(x)$, which is to move a sliding window with a fixed length L along $f(x)$ with a constant step size d . By centering the current data point in the middle of the sliding window, M frames of segments are created. In each frame, 10-level 1-D wavelet decomposition is adopted by selecting Daubechies-4 wavelets. As a result, a transformed version of $f(x)$ is created. If subject 1's FW resultant acceleration is taken as an example, a segment of original unlabeled FW resultant acceleration from 0 to 300 s is shown in Fig. 7(a). After extracting the WEE feature, it is transformed to the form in Fig. 7(b) in which zero crossings of the value of WEE occur. This leads to a difficulty of selecting the optimal threshold of endpoint detection. Thus, the absolute value of WEE is calculated as

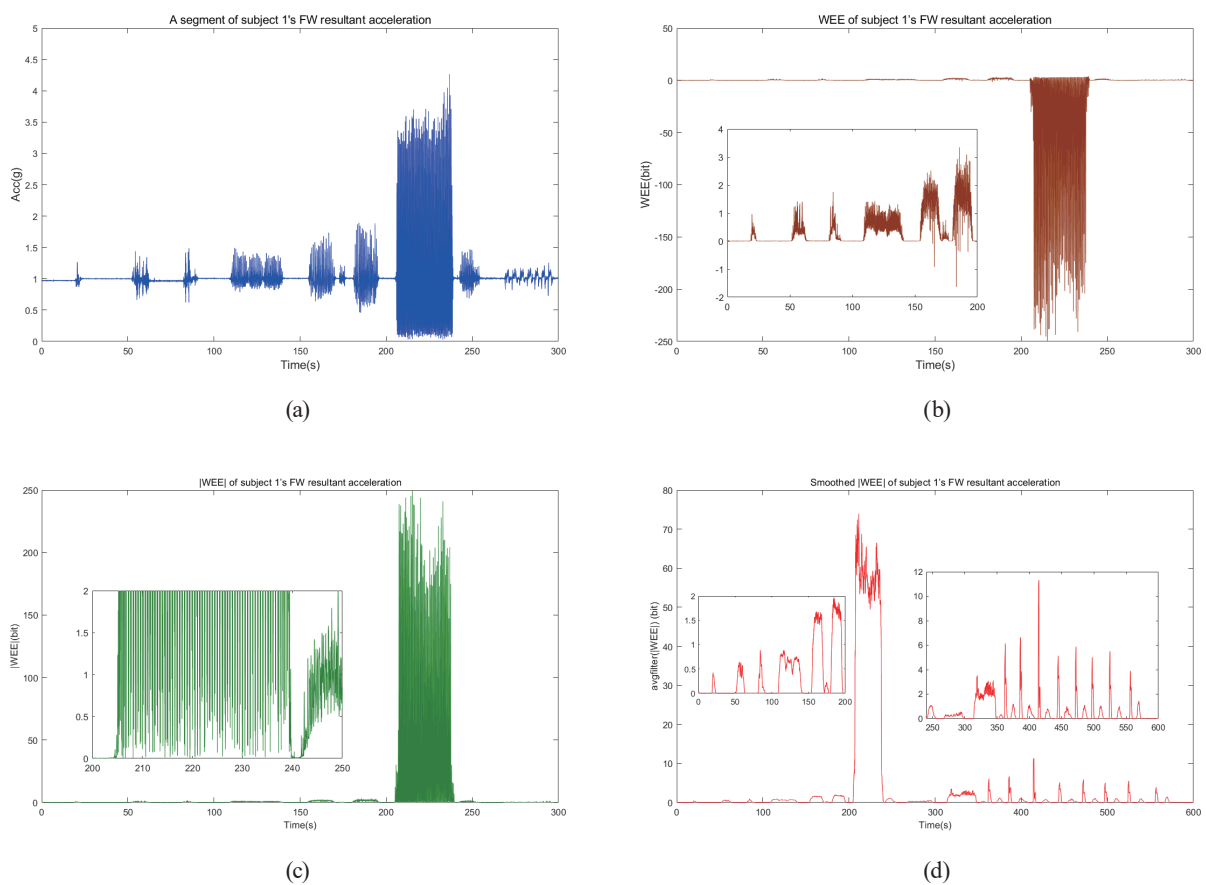


Fig. 7. (Color online) (a) A segment of subject 1's FW resultant acceleration, (b) WEE, (c) absolute WEE, and (d) smoothed absolute WEE of subject 1's FW resultant acceleration.

shown in Fig. 7(c) from which the zero-crossing rate of $|WEE|$ is lower than that in Fig. 7(b) qualitatively. However, from 200 to 240 s, a big group of data points with values approximating to zero still exists. To guarantee the quality of selected threshold for endpoint detection, average filtering was applied to smoothen $|WEE|$. Eventually, a soothing transformed version of $f(x)$ is obtained as shown in Fig. 7(d), which is taken as the base signal for automatic labeling in Sect. 3.3.

3.3 Automatic labeling

The original signal $f(x)$ is transformed into a smoothed absolute WEE of $f(x)$, which is denoted as $SA_WEE(x)$. In comparison with Fig. 7(a), the fact that $SA_WEE(x)$ is all positive and usually similar to a bell curve within activity intervals enables threshold-based endpoint detection. In this work, automatic labeling is divided into three main steps that include preliminary segmentation, endpoint detection, and assign labels.

In the preliminary segmentation stage, the primal goal is to find a suitable threshold that can distinguish activity intervals from stationary intervals in $SA_WEE(x)$. Activity intervals contain data points with larger values of $SA_WEE(x)$ than the selected threshold and indicate QPACT, SPACT, and some transition activities, i.e., SIT to STND, STND to LD, LD to STND, turning around before DS, WLK from location C to B as shown in Fig. 5 after finishing JUD, and LD to STND after each simulated fall. In contrast, stationary intervals contain data points with no larger values of $SA_WEE(x)$ than the selected threshold and indicate STACT and RP. As observed from collected activity data, e.g., as shown in Fig. 6, the duration of all activity/stationary intervals is approximately 50% of the total sampling time. Denote the sampling rate of each IMU as R_f , the total sampling time as T , the total number of data points in all activity intervals as N_a , and the total number of data points in all stationary intervals as N_s . Define activity rate as the proportion of N_a over the total number of data points in the total sampling time, which can be described as $R_a = N_a / (R_f \times T)$. Set the estimated activity rate ER_a to 50%. We summarized the method to find the initial segmenting threshold in the preliminary segmentation stage as shown in Algorithm 1. Figure 8(a) shows the result of the preliminary segmentation of subject 1's FW acceleration, which contains incorrectly segmented activity intervals with various durations. This issue is dealt with by the endpoint detection step.

The endpoint detection step is to determine the boundaries of activity intervals. Segmentation results from the preliminary segmentation stage are assigned with binary labels in which the activity intervals are labeled with one and the stationary intervals are labeled with zero. However, this binary series contains many incorrectly labeled intervals since the initial segmenting threshold is sensitive to local turbulence and transition activities. The strategy in this work to eliminate false segments is to extract constraints and information from HAS, which found the basis of an error correction of the binary series. According to HAS, the time stamps separating STACT from QPACT and QPACT from SPACT can be obtained as $T_1 = 100$ s and $T_2 = 340$ s, respectively. Define the upper bound of the shortest duration of an activity interval as $\delta_{dur} = 2.5$ s. A batch error correction is firstly done to set the label of any interval with a very short duration to zero. Transition activities create prominent acceleration turbulence that may affect the correct labeling in the preliminary segmentation and correspond

Algorithm 1: Preliminary Segmentation

Input: smoothed trend of absolute wavelet energy entropy, $SA_WEE(x)$; incremental step size for the amplification factor of the initial threshold value, $step_incre=0.01$; amplification factor to accelerate the optimal initial threshold searching, $amp_f=1$; estimated activity rate, $ER_a=0.5$; length of $SA_WEE(x)$, len ;

Output: Column indices of activity interval boundaries, col ;

```

1  target_vec=SA_WEE(x);
2  len=length(target_vec);
3  while 1 do
4    thresh=amp_f * mean(target_vec(1:500));
5    find col such that target_vec > thresh;
6    Ra=length(col)/len;
7    if |Ra - ERa| > step_incre do
8      amp_f ← amp_f + step_incre;
9    else
10   break;
11   end if
12 end while
13 return col;

```

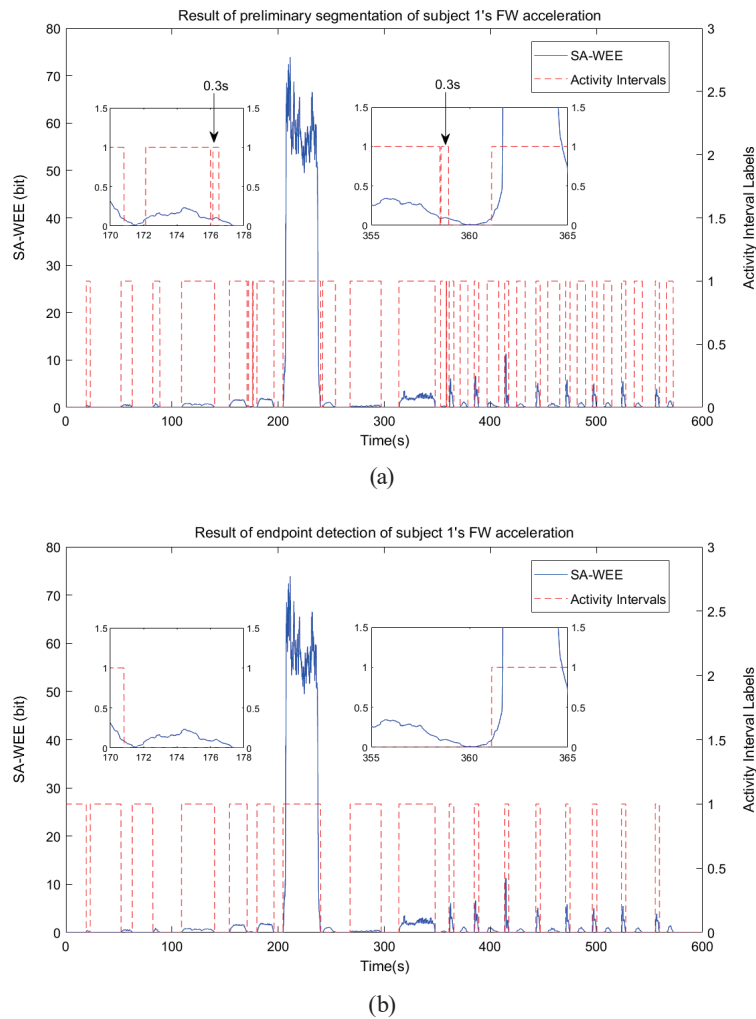


Fig. 8. (Color online) (a) Result of preliminary segmentation of subject 1's FW acceleration and (b) result of endpoint detection of subject 1's FW acceleration.

to the time-aligned bell curves on the curve of $SA_WEE(x)$. For instance, intervals containing transition activities between 0 and 100 s can be labeled as one, while zero is the label we expect for transition activities. Thus, label reversion needs to be done before the right boundary of the third transition activities (i.e., LD to STND) between 0 and 100 s. Denote the duration of a transition activity as δ_{tran} . As for QPACT between 100 and 340 s, transition activities occur between AS and DS, and after JUD. For the former transition activity, there is $\delta_{tran} \in (2.5 \text{ s}, 10 \text{ s}]$. For the latter transition activity, there is $\delta_{tran} \in (10 \text{ s}, 30 \text{ s}]$ since steady and slow moves are required during all transition activities. As for SPACT between 100 and 340 s, the upper bound of the duration of a complete fall is defined as $\delta_f = 6 \text{ s}$, and the lower bound of the maximum acceleration during a complete fall is defined as $\delta_{facc} = 1.8 \text{ g}$. On the basis of the abovementioned constraints and information, we summarized the algorithm for the endpoint detection stage as shown in Algorithm 2. Figure 8(b) shows the result of the endpoint detection of subject 1's FW acceleration. In contrast to Fig. 8(a), all incorrectly segmented activity intervals are corrected.

The final step is to label the result exported by the endpoint detection stage according to the order of target activities predefined by HAS, which is to assign decimal numbers from 1 to 17 to all activity intervals in sequence. Figure 9 shows the automatic labeling result of subject 1's FW acceleration in which all target activity intervals are correctly assigned with the corresponding labels from a qualitative perspective.

4. Experimental Results and Discussion

4.1 ALF performance indicators

The purpose of this study is to reduce manual efforts in labeling raw time series data regarding body-worn sensor-based HAR studies, which conduct data acquisition in a laboratory setting. Hence, two indicators, overall labeling accuracy (OLA) and average labeling time (ALT), are proposed to measure the performance of the proposed ALF. To obtain OLA, ground-truth labels should be given as a benchmark. During data acquisition, each subject's activities were recorded by a video camera, which helped annotators interpret and perform *post hoc* labeling on time series data. Moreover, to obtain a better insight into OLA, precision, recall, and *F*-measure are also adopted as subindicators of OLA. To reduce the manual annotating error, four annotators were asked to conduct data labeling, and the average result of activity endpoints was taken as the ground-truth labels of collected activity data. To obtain ALT, labeling time including label checking and correction time taken by each annotator during the whole manual labeling process was timed and then averaged. Note that all IMU modules were synchronized before data acquisition so that all IMU modules collected human activity data simultaneously. Thus, labeling raw data from all IMU modules is completed once labeling raw data from one module is carried out. We selected data from the FW IMU module as our target of automatic labeling. In addition, the proposed ALF was also examined on a modified dataset presented by Ref. 13. The performance of the proposed ALF was evaluated and verified using MATLAB R2016b, which was run in a Windows 10 $\times 64$ operating system with CoreTM i7-3612QM 2.10 GHz CPU and 8 GB memory.

Algorithm 2: Endpoint Detection

Input: data sampling rate, $R_f=100$; upper bound of the shortest duration of an activity interval, $\delta_{sdur}=2.5$; timestamp to separate STACT from QPACT, $T_1=100 * R_f$; timestamp to separate QPACT from SPACT, $T_2=340 * R_f$; upper bound of transition activity between AS and DS, $trans_1=10 * R_f$; upper bound of transition activity between JUD and SS, $trans_2=30 * R_f$; lower bound of the maximum acceleration during a complete fall, $\delta_{facc}=1.8$; upper bound of the duration of a complete fall, $\delta_f=6 * R_f$; Column indices of activity interval boundaries, col ;

Output: binary labels of the original time series, b_label ;

- 1 Create an all-one vector as b_label with the size of $1 \times len$
- 2 **for** each i in $(1 \leq i \leq length(col))$ **do**
- 3 $b_label(col(i)) = 0$;
- 4 **end for**
- 5 calculate each activity duration in b_label as len_zone
- 6 **for** each i in $(1 \leq i \leq length(len_zone))$ **do**
- 7 **if** $len_zone(i) < \delta_{sdur} * R_f$ **do**
- 8 **find** corresponding segments in b_label **do** set labels to 0
- 9 **end if**
- 10 **end for**
- 11 **update** b_label ;
- 12 $ts = [1 \ T_1 \ T_2 \ len]$;
- 13 **while** each n in $(1 \leq n \leq 3)$ **do**
- 14 **if** $n == 1$ **do**
- 15 calculate the number of segments in $b_label(ts(n):ts(n+1))$ and assign it to seg_n
- 16 **if** $seg_n == 7$ **do**
- 17 **for** each i in $(1 \leq i \leq T_1)$ **do**
- 18 $b_label(i) \leftarrow b_label(i) + 1$;
- 19 **if** $b_label(i) == 2$ **do**
- 20 $b_label(i) = 0$;
- 21 **end if**
- 22 **end for**
- 23 $n \leftarrow n + 1$
- 24 continue;
- 25 **else do**
- 26 merge proximal segments with identical labels and **update** b_label ;
- 27 **end if**
- 28 **else if** $n == 2$ **do**
- 29 calculate each activity duration in $b_label(ts(n):ts(n+1))$ as len_zone ;
- 30 **find** segments such that $\delta_{sdur} < len_zone \leq trans_1$ **do** set labels to 0;
- 31 **find** segments such that $trans_1 < len_zone \leq trans_2$ **do** set labels to 0;
- 32 $n \leftarrow n + 1$;
- 33 continue;
- 34 **else if** $n == 3$ **do**
- 35 calculate each activity duration in $b_label(ts(n):ts(n+1))$ as len_zone ;
- 36 calculate the maximum value in each activity interval as m_acc ;
- 37 **find** segments such that $len_zone > \delta_f$ **or** $m_acc < \delta_{facc}$ **do** set labels to 0;
- 38 $n \leftarrow n + 1$;
- 39 **end if**
- 40 **end while**
- 41 **return** b_label

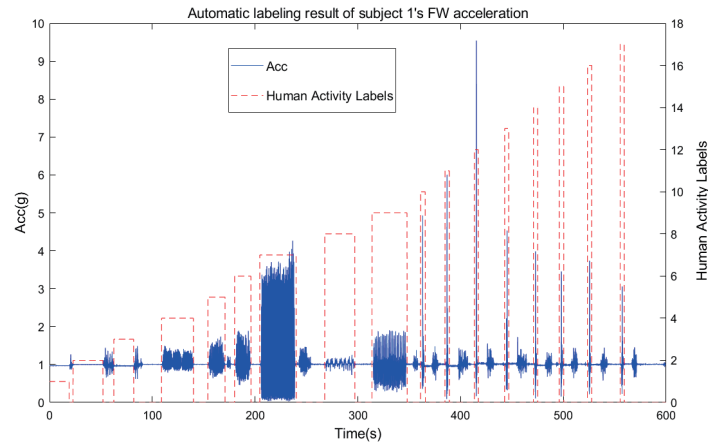


Fig. 9. (Color online) Automatic labeling result of subject 1's FW acceleration.

4.2 OLA

The OLA is defined as the average ratio per subject of the number of correctly labeled data points over the total number of data points as shown in Eq. (6).

$$R_{OLA} = \frac{1}{6} \sum_{i=1}^6 \frac{N_c^{(i)}}{N_{all}^{(i)}}, \quad (6)$$

where $N_c^{(i)}$ is the number of correctly labeled data points and $N_{all}^{(i)}$ is the total number of data points of the subject i . Comparison between labels created by the proposed ALF and the ground truth was made to obtain R_{OLA} . The labeling accuracies of each subject's activity data are obtained as 96.2, 94.5, 95.6, 96.2, 96.4, and 96.1% from subjects 1 to 6, respectively. Thus, R_{OLA} is 95.8%. If the automatic labeling result is taken as the response of ALF in a supervised learning perspective, F -measure can be taken as a performance measurement. Precision, recall, and F -measure are usually used in a binary classification setting and consist of three scores including true positive (TP), false positive (FP), and false negative (FN).⁽²⁶⁾ In our work, nonzero labels (i.e., 17 target activities) are treated as positive, and zero labels (i.e., transition activities and RP) are treated as negative. In this manner, the multilabel labeling issue in this work is turned into a binary labeling issue. The results of precision, recall, F -measure, and accuracy of the proposed ALF tested using collected activity data are presented in Table 1 in which 'M' denotes a male subject and 'F' denotes a female subject on the row of subject number. The average precision, recall, and F -measure are 91.0, 99.5, and 95.0%, respectively. For intervals being automatically assigned with corresponding target activity labels, 91.0% of the labels are correctly assigned, and 99.5% of true target activities are correctly assigned with the corresponding labels.

Table 1
Precision, recall, F-measure, and accuracy of ALF on collected data.

Subject No.	1(M)	2(F)	3(M)	4(M)	5(M)	6(M)	Average
Precision (%)	91.5	88.6	90.2	91.8	92.4	91.3	91.0
Recall (%)	99.7	98.7	99.5	99.6	99.6	99.7	99.5
F-measure (%)	95.4	93.4	94.7	95.5	95.9	95.3	95.0
Accuracy (%)	96.2	94.5	95.6	96.2	96.4	96.1	95.8

In Table 1, subject 2's precision is lower than 90.0%. Further comparison analysis between automatic labeling results and ground truth shows that the durations of the longest mislabeled segments of each subject's data are 1.15, 6.19, 2.86, 1.45, 1.16, and 1.46 s. Figures 10(a) and 10(b) show the results of automatic labeled and mislabeled intervals of subject 2's and 3's FW accelerations. Mislabeled intervals are assigned with a negative label of -5 . In both situations, transition activities (marked in purple dashed circles) that occurred at the beginning or end of a target activity are incorrectly assigned with the next or former target activity label. Even though all subjects were instructed in advance that a stationary status (i.e., staying in RP) is necessary after and before performing a target activity, some degree of inaccuracy in performing each target activity can still happen. Figures 10(c)–10(f) show the remaining results of automatic labeled and mislabeled intervals of other subjects' FW accelerations.

In Ref. 13, a smartphone was used to collect six activities, namely, walking, walking upstairs, walking downstairs, sitting, standing, and laying, and contributed it as the UCI HAR Dataset (UHD). However, the RP status proposed in this work was not incorporated into the data acquisition process in Ref. 13. Thus, a slight modification of the UCI HAR Dataset was carried out so that a performance evaluation on the proposed ALF can be conducted on another public dataset. During data acquisition of Ref. 13, subjects were instructed to perform each activity freely at least twice in a predefined sequence, i.e., standing (30 s) \rightarrow sitting (30 s) \rightarrow lying down (30 s) \rightarrow walking (30 s) \rightarrow walking downstairs (36 s) \rightarrow walking upstairs (36 s), and each sequence of activity was performed twice. Thus, RPs were inserted between each target activity, and a HAS of a modified UCI HAR Dataset (mUHD-HAS) was created as shown in Fig. 11. Particularly, to maintain the structure of the original dataset, RP data was created by sampling stationary activities in the current time series, i.e., RP inserted after STND is sampled from the last 10 s of the previous STND, RP inserted after SIT is sampled from the last 10 s of the previous SIT, RP inserted after LD is sampled from the last 10 s of the previous STND owing to the fact that one needs to get up before walking, and RPs inserted between WLK and DS and between DS and AS are both sampled from the first 10 s of the previous STND owing to the fact that one would be standing as an RP. In this way, data from 30 volunteers in the UCI HAR Dataset were modified as shown in Fig. 11 and tested by adopting the proposed ALF. Figures 12(a) and 12(b) show subject 2's resultant acceleration from the UCI HAR Dataset before and after modification and the corresponding original labels.

Before applying ALF to the modified UCI HAR Dataset, constraints and information were extracted from mUHD-HAS. Set the estimated activity rate ER_a to 40%. Assume that the

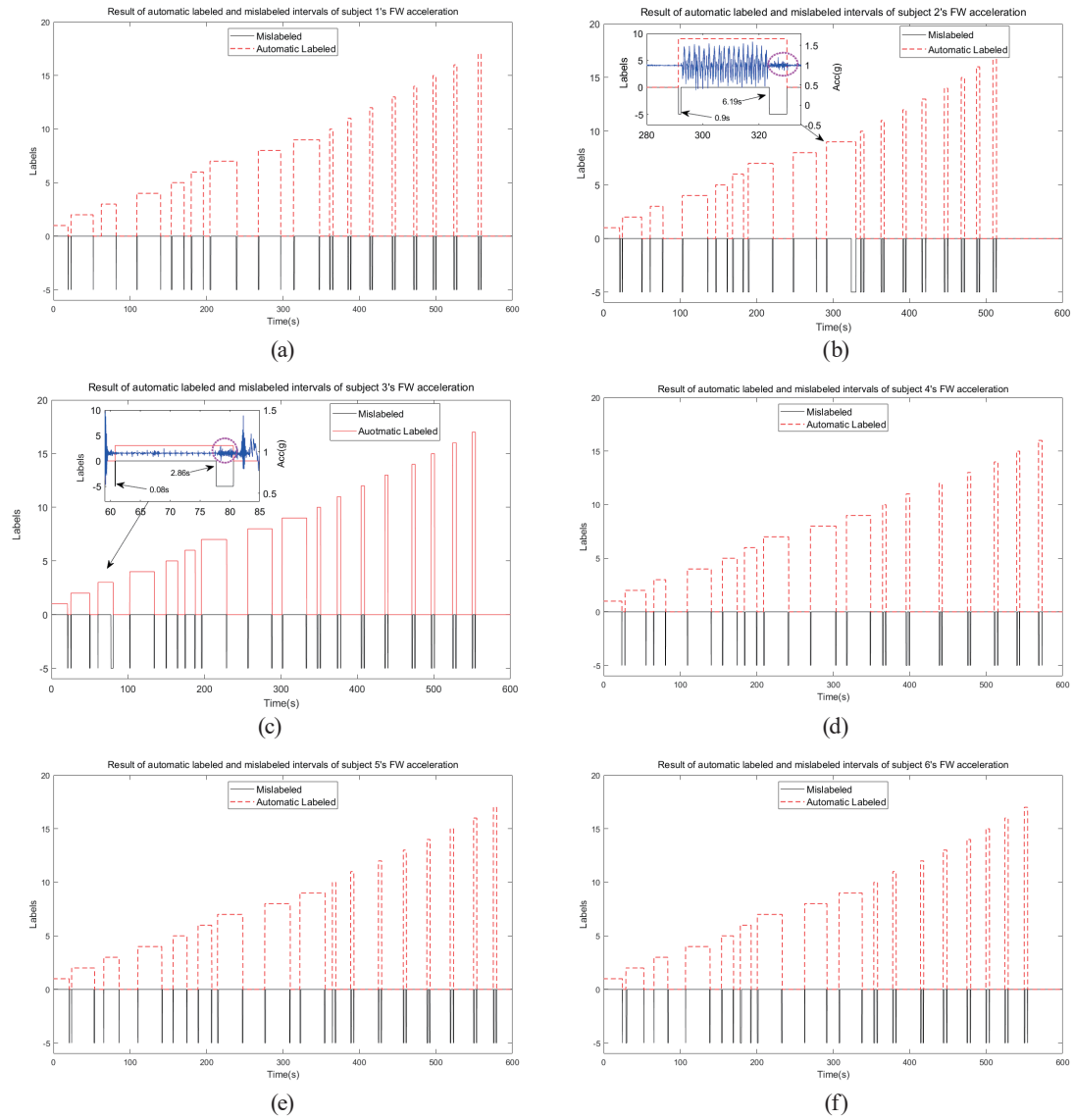


Fig. 10. (Color online) (a) Results of automatic labeled and mislabeled intervals of subject 1's, (b) subject 2's, (c) subject 3's, (d) subject 4's, (e) subject 5's, and (f) subject 6's FW accelerations.

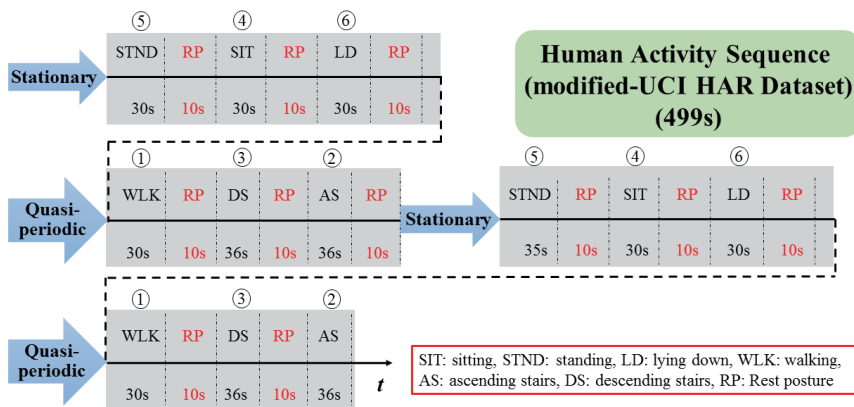


Fig. 11. (Color online) HAS of modified UCI HAR Dataset.

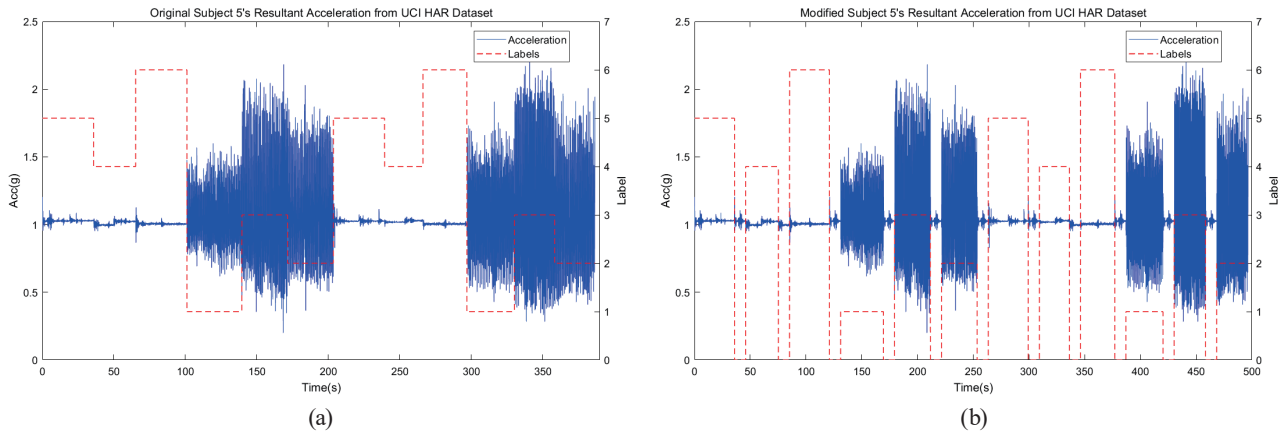


Fig. 12. (Color online) (a) Original subject 5's resultant acceleration and (b) modified subject 5's resultant acceleration from UCI HAR Dataset.

duration of the current subject's activity data is T . Since the duration of data collected from each subject varies from each other in the original UHD, the time stamps separating stationary activities from QPACT were obtained as $T_1 = 120 \times T/499$, $T_2 = 252 \times T/499$, and $T_3 = 377 \times T/499$ according to mUHD-HAS. Choose the upper bound of the shortest duration of an activity interval as $\delta_{sdur} = 5$ s. Since the modification of UHD might introduce acceleration turbulence into the RP stage, we chose the duration of a transition activity as $\delta_{tran} \in [0, 10]$ s. Particularly for stationary activities between 0 s and T_1 , and between T_2 and T_3 , the strategy of assigning labels to each stationary activity was changed to first divide both intervals $[0, T_1]$ and $[T_2, T_3]$ into three intervals with identical durations and then to assign the corresponding labels to each interval according to mUHD-HAS. On the basis of the abovementioned constraints and information, precision, recall, F -measure, and accuracy were adopted to evaluate the performance of the proposed ALF on 18 volunteers' activity data from mUHD. The remaining 12 volunteers' activity data (i.e., subjects 1, 9, 17, 18, 19, 21, 22, 23, 24, 26, 28, and 30) are excluded from the test since none of the sequences of their collected data follows the mUHD-HAS. Intervals with a zero label are treated as negative samples while others with nonzero labels are treated as positive samples. Results are presented in Table 2. In Table 2, the average accuracy is 82.1% and the average precision is 81.9%. This may be caused by the inaccuracy of choosing the separating timestamps, which is a consequence of the uncertainty in the durations of subjects' data. The average recall is 97.3%, which indicates that the majority of target activities are correctly labeled, and the corresponding average F -measure is 88.9%.

One limitation of this work is that errors are propagated after preliminary segmentation because the preliminary segmentation threshold is chosen according to the estimated activity rate that is subject to the design of HAS and the observation of the collected data, which makes the aforementioned batch error correction after preliminary segmentation necessary. The results of the proposed ALF on the collected dataset [Figs. 10(a)–10(f)] and the mUHD show that activity intervals detected by ALF are generally wider than those defined by ground truth labels, which leads to a certain number of FPs/FNs regarding each target activity. As the

Table 2
Precision, recall, F -measure, and accuracy of ALF on modified UCI HAR Dataset.

Subject No.	Precision (%)	Recall (%)	F -measure (%)	Accuracy (%)
2	82.4	99.4	90.1	83.8
3	83.7	99.1	90.7	84.6
4	81.9	99.7	90.0	83.5
5	79.1	99.0	87.9	80.6
6	82.6	98.8	90.0	83.5
7	82.1	99.2	89.9	83.4
8	81.6	96.8	88.5	81.6
10	81.7	99.0	89.5	83.1
11	83.3	99.3	90.6	84.5
12	82.1	97.7	89.2	82.5
13	79.2	97.7	87.5	79.8
14	84.0	100.0	91.3	85.6
15	83.1	96.1	89.2	82.5
16	80.9	92.2	86.1	78.0
20	77.4	91.5	83.9	74.6
25	87.6	99.3	93.1	88.3
27	81.0	91.8	86.0	77.7
29	81.1	94.4	87.2	79.4
Average	81.9	97.3	88.9	82.1

results show, for SPACT, activity intervals defined by ground truth labels are enclosed by those detected by ALF, which means that there are only FPs and no FN. For STACT and QPACT, there are both FPs and FNs. Since a fall could be fatal to seniors, the risk of having more FNs than FPs after adopting ALF is higher to SPACT than to STACT and QPACT. A possible strategy to lower such risk is to select a smaller initial threshold, namely, a larger ER_a , for preliminary segmentation, which expands the activity intervals but increases the error rate of preliminary segmentation.

4.3 ALT

Each annotator labeled activity data collected from six subjects all at once. The labeling time of each annotator is presented in Table 3. The ALT is 76.8 min. However, the automatic labeling time of the proposed ALF is 1116.71 s, which is 18.6 min in total. Therefore, the time spent on automatic labeling using the proposed ALF is 75.8% less than the average time spent on manual labeling. In addition, the automatic labeling time of the proposed ALF on the modified UCI HAR Dataset (18 subjects) is 2160.296 s, which is approximately 36.0 min in total. Since there is no information about the time spent on labeling the whole dataset from Ref. 13 and because of the lack of complete video recordings of the UCI HAR Dataset, no comparison is done between the manual labeling time and the automatic labeling time using ALF on the UCI HAR Dataset.

Table 3
Labeling time of ALF on collected data.

Annotator No.	1 (min)	2 (min)	3 (min)	4 (min)	Average/Total (min)
Manual labeling time	81	78	80	68	76.8
Automatic labeling time	—	—	—	—	18.6

5. Conclusions

HAR has become a highlighted research area over the last few years since related research outputs play an increasingly important role in our aging society. In this study, we aimed at reducing labeling efforts on time series data, which is collected over diverse individuals using multiple body-worn IMUs in a laboratory setting. Instead of using a small amount of labeled data to gain a robust classifier, this work focuses on developing an ALF to directly assign accurate labels to unlabeled raw data. In the proposed ALF, a HAS is the information center that is predefined as a sequence of different target activities concatenated by RPs providing constraints to improve the robustness of an automatic labeling algorithm. This algorithm, as the execution part, firstly transformed the original activity data into its corresponding trend of absolute WEE, then segmented it into activity intervals and stationary intervals based on constraints and information extracted from HAS. From the experimental results, an OLA of 95.8% was obtained with the average precision, average recall, and average F-measure score as 91.0, 99.5, and 95.0%, respectively, when the proposed ALF was tested on our collected data. The total labeling time for the proposed framework is approximately 18.6 min, which shortens the manual labeling time (average of 76.8 min) by 75.8%. A public dataset, the UCI HAR Dataset, was modified to the proposed ALF. We obtained the average precision of 81.9% and the overall accuracy of 82.1%, and the average recall and the average F-measure are 97.3% and 88.9%, respectively. The total time for automatically labeling the modified UCI HAR Dataset (18 subjects) is approximately 36.0 min. Both experimental results showed that the proposed ALF can reduce the labeling efforts significantly with a guarantee of labeling accuracy, and the framework can be adopted as a rapid and reliable way of generating labeled datasets.

Acknowledgments

This research is supported by the National Natural Science Foundation for Young Scientists of China (Grant No. 61501034). We also thank Shanda Wang for useful discussions and all our colleagues for providing all types of help during the preparation of this manuscript.

References

- 1 A. Bulling, U. Blanke, and B. Schiele: ACM Comput. Surv. (CSUR) **46** (2014) 1. <https://dx.doi.org/10.1145/2499621>
- 2 K. V. Laerhoven, D. Kilian, and B. Schiele: IEEE Int. Symp. Wearable Computers (2008) 63. <https://dx.doi.org/10.1109/ISWC.2008.4911586>
- 3 E. M. Tapia, S. S. Intille, and K. Larson: Int. Conf. Pervasive Computing (2004) 158. <http://courses.media.mit.edu/2004fall/mas622j/04.projects/home/TapiaIntilleLarson04.pdf>

- 4 B. Logan, J. Healey, M. Philipose, E. M. Tapia, and S. Intille: UBIComp 2007: Ubiquitous Computing, Int. Conf., UBIComp 2007, Innsbruck, Austria, September 16–19, 2007, Proc. (2008) 483. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.137.1168&rep=rep1&type=pdf>
- 5 U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher: Int. Workshop Wearable and Implantable Body Sensor Networks (2006) 113. <https://dx.doi.org/10.1109/BSN.2006.6>
- 6 M. Fritz and B. Schiele: Int. Conf. Ubiquitous Computing (2008) 10. <https://dx.doi.org/10.1145/1409635.1409638>
- 7 J. Froehlich, M. Y. Chen, S. Consolvo, B. Harrison, and J. A. Landay: Int. Conf. Mobile Systems, Applications and Services (2007) 57. <https://dx.doi.org/10.1145/1247660.1247670>
- 8 A. Bulling, J. A. Ward, and H. Gellersen: ACM Trans. Appl. Percept. **9** (2012) 1. <https://dx.doi.org/10.1145/2134203.2134205>
- 9 D. Roggen, K. Forster, A. Calatroni, and T. Holleczeck: World of Wireless, Mobile and Multimedia Networks & Workshops, 2009, WoWMoM 2009, 10th IEEE Int. Symp. (2009) 1. <https://dx.doi.org/10.1109/WOWMOM.2009.5282442>
- 10 U. Blanke and B. Schiele: Int. Symp. Wearable Computers (ISWC) (2010) 1. <https://dx.doi.org/10.1109/ISWC.2010.5665869>
- 11 M. Stikic, D. Larlus, S. Ebert, and B. Schiele: IEEE Trans. Pattern Anal. Mach. Intell. **33** (2011) 2521. <https://dx.doi.org/10.1109/TPAMI.2011.36>
- 12 M. Plotnik, D. Roggen, I. Muidan, J. M. Hausdorff, and N. Giladi: IEEE Trans. Inf. Technol. Biomed. **14** (2010) 436. <https://dx.doi.org/10.1109/TITB.2009.2036165>
- 13 D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz: ESANN 2013, 21st European Symp. Artificial Neural Networks, Computational Intelligence and Machine Learning (2013) 437. <https://www.elen.ucl.ac.be/Proc/esann/esannpdf/es2013-84.pdf>
- 14 O. Banos, R. Garcia, J. A. Holgado-Terriza, M. Damas, H. Pomares, I. Rojas, A. Saez, and C. Villalonga: 6th Int. Work-Confer., IWAAL 2014 LNCS (2014) 91. https://dx.doi.org/10.1007/978-3-319-13105-4_14
- 15 K. Davis, E. Owusu, V. Bastani, and L. Marcenaro: Int. Conf. Information Fusion (2016) 371. <https://ieeexplore.ieee.org/document/7527912/>
- 16 B. Bruno, F. Mastrogiovanni, A. Sgorbissa, and T. Vernazza: IEEE Int. Conf. Automation Science and Engineering (2012) 156. <https://dx.doi.org/10.1109/CoASE.2012.6386410>
- 17 B. Bruno, F. Mastrogiovanni, A. Sgorbissa, and T. Vernazza: IEEE Int. Conf. Robotics and Automation (2013) 1602. <https://dx.doi.org/10.1109/ICRA.2013.6630784>
- 18 K. Altun, B. Barshan, and O. Tunçel: Pattern Recognit. **43** (2010) 3605. <https://dx.doi.org/10.1016/j.patcog.2010.04.019>
- 19 X. Guan, R. Raich, and W. K. Wong: Int. Conf. Machine Learning (2016) 2330. <http://proceedings.mlr.press/v48/guan16.pdf>
- 20 D. Wyatt, M. Philipose, and T. Choudhury: The Twentieth National Conf. Artificial Intelligence and the Seventeenth Innovative Applications of Artificial Intelligence Conf., July 9–13, 2005, Pittsburgh, Pennsylvania, USA (2005) 21. <https://www.aaai.org/Papers/AAAI/2005/AAAI05-004.pdf>
- 21 S. Böttcher, P. M. Scholl, and K. Van Laerhoven: Int. Workshop on Sensor-Based Activity Recognition and Interaction (2017) 1. <https://dx.doi.org/10.1145/3134230.3134233>
- 22 J. P. Byrnes: Cognitive Development and Learning in Instructional Contexts (Allyn and Bacon, Boston, 2001). <https://nojh49ddm01.storage.googleapis.com/MDIwNTUwNzcxOQ==01.pdf>
- 23 D. Cook, K. D. Feuz, and N. C. Krishnan: Knowledge Inf. Syst. **36** (2013) 537. <https://dx.doi.org/10.1007/s10115-013-0665-3>
- 24 T. L. M. V. Kasteren, G. Englebienne, and B. J. A. Kröse: the 2008 AAAI Fall Symp. AI in Eldercare: New Solutions to Old Problems FS-08-02 (2008) 142. <https://www.aaai.org/Papers/Symposia/Fall/2008/FS-08-02/FS08-02-023.pdf>
- 25 T. L. M. V. Kasteren, G. Englebienne, and B. J. A. Kröse: 8th Int. Conf., Pervasive 2010 **6030** (2010) 283. https://dx.doi.org/10.1007/978-3-642-12654-3_17
- 26 W. H. Chen, P. C. Cho, and Y. L. Jiang: Sens. Mater. **29** (2017) 897. <https://dx.doi.org/10.18494/SAM.2017.1546>
- 27 O. Chapelle, B. Schölkopf, and A. Zien: Semi-Supervised Text Classification Using EM (Cambridge, Mass., MIT Press, 2006) pp. 33–55. http://parsec.nuaa.edu.cn/seminar/2012_Spring/20120323/%E8%92%8B%E8%90%8D/Semi-Supervised%20Text%20Classification%20Using%20EM.pdf
- 28 M. Stikic, K. Van Laerhoven, and B. Schiele: IEEE Int. Symp. Wearable Computers (2008) 81. <https://dx.doi.org/10.1109/ISWC.2008.4911590>
- 29 D. Guan, W. Yuan, Y. K. Lee, A. Gavrilov, and S. Lee: IEEE Int. Conf. Embedded and Real-Time Computing Systems and Applications (2007) 469. <https://dx.doi.org/10.1109/RTCSA.2007.17>
- 30 L. Zhao, G. Sukthankar, and R. Sukthankar: AAAI Conf. Human Computation (2011) 74. <https://www.aaai.org/ocs/index.php/WS/AAAIW11/paper/view/3874>

- 31 C. Lee and Y. Xu: IEEE Int. Conf. Robotics and Automation, 1996, Proc. (2002) 2982. <https://dx.doi.org/10.1109/ROBOT.1996.509165>
- 32 O. Amft, H. Junker, and G. Troster: IEEE Int. Symp. Wearable Computers, 2005, Proc. (2005) 160. <https://dx.doi.org/10.1109/ISWC.2005.17>
- 33 M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. Havinga: Sensors **14** (2014) 10146. <https://dx.doi.org/10.3390/s140610146>
- 34 D. Riboni: Pers. Ubiquitous Comput. **15** (2011) 271. <https://dx.doi.org/10.1007/s00779-010-0331-7>
- 35 S. Patel, C. Mancinelli, P. Bonato, J. Healey, and M. Moy: 2009 Sixth Int. Workshop on Wearable and Implantable Body Sensor Networks (2009) 234. <https://dx.doi.org/10.1109/BSN.2009.53>
- 36 J. K. Min and S. B. Cho: IEEE Int. Conf. Systems, Man, and Cybernetics (2011) 1319. <https://dx.doi.org/10.1109/ICSMC.2011.6083808>
- 37 S. S. Intille, F. Albinali, S. Mota, and B. Kuris: Engineering in Medicine and Biology Society, Embe, 2011 Int. Conf. IEEE (2011) 3636. <https://dx.doi.org/10.1109/IEMBS.2011.6090611>
- 38 J. Carter, D. Wilkinson, S. Blacker, M. Rayson, J. Bilzon, R. Izard, A. Coward, A. Wright, A. Nevill, and K. Rennie: J. Sports Sci. **26** (2008) 553. <https://dx.doi.org/10.1080/02640410701708979>
- 39 V. T. van Hees, F. Renström, A. Wright, A. Gradmark, M. Catt, K. Y. Chen, M. Löf, L. Bluck, J. Pomeroy, and N. J. Wareham: PLoS One **6** (2011) e22922. <https://dx.doi.org/10.1371/journal.pone.0022922>
- 40 N. Pannurat, S. Thiemjarus, E. Nantajeewarawat, and I. Anantavasilp: Sensors **17** (2017) 774. <https://dx.doi.org/10.3390/s17040774>
- 41 L. S. Huang and C. H. Yang: IEEE Int. Conf. Acoustics, Speech, and Signal Processing **3** (2000) 1751. <https://dx.doi.org/10.1109/ICASSP.2000.862091>
- 42 C. Jia and B. Xu: Int. Symp. Chinese Spoken Language Processing (ISCSLP 2002) (2002) 96. https://isca-speech.org/archive_open/archive_papers/isclsp2002/clp2_096.pdf
- 43 W. H. Shin, B. S. Lee, Y. K. Lee, and J. S. Lee: IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proc. (2000) 1399. <https://dx.doi.org/10.1109/ICASSP.2000.861845>
- 44 J. L. Shen, J. W. Hung, and L. S. Lee: Int. Conf. Spoken Language Processing, Incorporating Australian Int. Speech Science and Technology Conf. (1998) 0232. https://www.isca-speech.org/archive/archive_papers/icslp_1998/i98_0232.pdf