Journal of Orthopaedic Translation (2017) 11, 19—29

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: http://ees.elsevier.com/jot

ORIGINAL ARTICLE

Clinical validation of a body-fixed 3D accelerometer and algorithm for activity monitoring in orthopaedic patients

Matthijs Lipperts [a,b,*], Simon van Laarhoven [b], Rachel Senden [b], Ide Heyligers [b], Bernd Grimm [b]

[a] *School for Medical Physics and Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*
[b] *Atrium Medical Centre Heerlen Orthopaedic Research and Scientific Education, Department of Orthopaedics, Zuyderland Medical Center, Heerlen, The Netherlands*

Received 30 November 2016; received in revised form 15 January 2017; accepted 6 February 2017
Available online 27 February 2017

KEYWORDS
accelerometry;
activity monitoring;
clinical outcome assessment

Summary *Background/Objective:* Activity is increasingly being recognized as a highly relevant parameter in all areas of healthcare for diagnosis, treatment, or outcome assessment, especially in orthopaedics where the movement apparatus is directly affected. Therefore, the aim of this study was to develop, describe, and clinically validate a generic activity-monitoring algorithm, satisfying a combination of three criteria. The algorithm must be able to identify, count, and time a large set of relevant daily activities. It must be validated for orthopaedic patients as well as healthy individuals, and the validation must be in a setting that mimics free-living conditions.
Methods: Using various technical solutions, such as a dual-axis approach, dynamic inclinometry (hip flexion), and semiautomatic calibration (gait speed), the algorithms were designed to count and time the following postures, transfers, and activities of daily living: resting/sitting, standing, walking, ascending and descending stairs, sit—stand transitions, and cycling. In addition, the number of steps per walking bout was determined. Validation was performed with healthy individuals and patients who had undergone unilateral total joint arthroplasty, representing a wide spectrum of functional capacity. Video observation was used as the gold standard to count and time activities in a validation protocol approaching free-living conditions.
Results: In total 992 and 390 events (activities or postures) were recorded in the healthy group and patient group, respectively. The mean error varied between 0% and 2.8% for the healthy group and between 0% and 7.5% for the patient group. The error expressed in percentage of time varied between 2.0% and 3.0% for both groups.
Conclusion: Activity monitoring of orthopaedic patients by counting and timing a large set of relevant daily life events is feasible in a user- and patient-friendly way and at high clinical

* Corresponding author. Laan van Hövell tot Westerflier 27, 6411EW Heerlen, The Netherlands.
E-mail address: matthijslipperts@gmail.com (M. Lipperts).

http://dx.doi.org/10.1016/j.jot.2017.02.003

validity using a generic three-dimensional accelerometer and algorithms based on empirical and physical methods. The algorithms performed well for healthy individuals as well as patients recovering after total joint replacement in a challenging validation set-up. With such a simple and transparent method real-life activity parameters can be collected in orthopaedic practice for diagnostics, treatments, outcome assessment, or biofeedback.

Introduction

Physical activity (PA) is increasingly recognized as a major contributor to general health [1] and thus rising efforts are being made to assess activity as a quantitative parameter in the medical field [1,2]. In orthopaedics, where the movement apparatus is directly affected and treated, assessing daily life activity as an outcome dimension independent of the commonly used questionnaire-based scores for pain, satisfaction, or function is of particular interest. For assessing PA some self-report questionnaires are available (Short QUestionnaire to ASsess Health-enhancing physical activity, Longitudinal Aging Study Amsterdam Physical Activity Questionnaire). However, in a review of 17 of such tools, including the most popular ones, none could meet clinimetric standards and consequently the use of accelerometers was advised for monitoring [3]. Also, a recently published recommendation of the Osteoarthritis Research Society International advised sensor-based activity monitoring (AM) to assess outcomes in patients with osteoarthritis [4].

Modern developments in sensor technology, such as miniaturization, have enabled the use of wearable sensor-based AM. So far, the focus of AM has mainly been on energy expenditure, which is especially popular to assess the effect of lifestyle and general health interventions for cardiac and pulmonary diseases and obesity [5]. Many accelerometer-based AM methods only provide activity counts or caloric expenditure [6,7] based on intensity count thresholds and caloric maps, instead of identifying, counting, and timing the actual activity events such as walking. In several medical fields, however, especially in orthopaedics, one is interested in the identification of specific motor tasks and counting and timing these well-defined events instead of finding the overall intensity or caloric burn. In a recent review [8] (on AM studies under free-living conditions in orthopaedic patients) it was shown that studies which used general quantitative activity parameters such as energy expenditure, time upright, or daily steps seemed less discriminative and responsive in orthopaedic applications while more specific event counts such as minutes of moderate and vigorous PA or climbing stairs were clinimetrically more powerful.

The goal of orthopaedic interventions, besides pain relief, is the restoration of musculoskeletal function to enable the performance of activities desired by the patient or required to live independently, to participate in society, and achieve a healthy lifestyle. Thus it is highly relevant in orthopaedic outcome assessment to investigate whether, when, how often, and for how long patients are able to perform relevant and possibly challenging activities of daily living (ADL), e.g., sitting, standing, sit—stand transitions, walking, cycling, and stair climbing and descending.

As an example for choosing activity events relevant for classification in orthopaedics, ascending stairs seems very appropriate as it is the energetically more demanding task [9]. However, it is conceivable that descending stairs is motorically more difficult for patients undergoing total joint arthroplasty because of pain, loss of muscle strength, joint instability, or proprioception and a fear of falling. Thus, in patients with lower limb osteoarthritis, counting and timing of both stair events should be a highly relevant outcome measure. Cycling, stationary or on a normal bike, is a common and often recommended or prescribed physiotherapeutic activity for orthopaedic patients recovering from surgery or the elderly osteoarthritic patient in general [10]. Cycling is also an important activity for social participation for many, so that its classification adds great value to outcome assessment or supervising the compliance to therapy. Step counters and most commercial monitors cannot distinguish between walking and cycling.

Several AM devices with analysis software such as Actigraph (Actigraphcorp, Pensacola, FL, USA), StepWatch (Modus Health, Washington, DC, USA), Shimmer3 (Shimmer, Dublin, Ireland), Dynaport (McRoberts, Den Haag, The Netherlands), ActivPal (Pal Technologies, Glasgow, UK), Physilog (Gait Up, Lausanne, Switzerland), RT3 (StayHealthy, Monrovia, CA, USA) and others have already been developed and are commercially available [7,8,11—17]. However, their algorithms are usually proprietary and nondisclosed, they do not identify all of the aforementioned activities, or are not always patient and user friendly (e.g., bulky). Furthermore, to date only a few studies have validated their algorithms on patients whose movement apparatus has been affected [18—20]. In these patients there is a broad range of activity levels, ranging from being able to walk only very short bouts with the help of walking aids (1st week postoperatively), to uninhibited movement at the level of a healthy individual. Not only does a patient's condition influence his or her activity level, but also the way a movement is performed (i.e., slower, lower intensity, use of walking aids) challenges the universal validity of signal analysis algorithms. Thus, this could affect the performance of AM devices when used to monitor orthopaedic patients [21—23]. Therefore, it is important to validate AM algorithms using individuals representing the intended target group.

Due to miniaturization of sensors and chips, it has become feasible to increase the data-storage capacity of devices, enabling 100% postprocessing of data. This in turn enables the creation of AM algorithms that are hardware

independent, and can be used with any accelerometer device, to allow for a widespread use of generic accelerometers without the need to buy expensive preprogramed devices.

The aim of this study was to develop, describe, and clinically validate a generic AM algorithm, satisfying a combination of three criteria. First of all the algorithm must be able to identify, count, and time a large set of relevant ADL as events. Second, it must be validated for orthopaedic patients as well as healthy individuals, and finally the validation must be in a setting mimicking free-living conditions. The innovative character of this study lies in the combination of these criteria.

Methods

All procedures performed were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Ethical approval was obtained from the 'Medisch-Ethische Toetsings Commissie Atrium-Orbis-Zuyd' under reference number: 10-N-72. Informed consent was obtained from all individual participants included in the study.

Device description

A commercially available three-dimensional accelerometer device (GC Dataconcepts, Waveland, MS, USA) was used to acquire data. The dimensions of this light-weight (18 g) device were 64 mm $\times$ 25 mm $\times$ 13 mm. Acceleration data (from $-6g$ to $+6g$) was sampled at 50 Hz and stored on an on-board memory chip with a capacity up to 16 GB and a battery life of $>$ 24 hours. Data transfer and device configuration were possible via USB.

Device location and positioning obviously influences the accelerometer signal, thus location and positioning is crucial in any AM study. Only one group developed position-independent algorithms [24]; however, this limits the activity types which can be classified. For example, they do not discriminate between sitting and standing and thus cannot count transfers. In this study a position on the lateral side of the nonaffected upper leg was chosen (Figure 1). This way the inclinometer function of an accelerometer during static periods with reference to earth gravity, as described by Godfrey et al [25], could be used to best advantage. Direct attachment of the sensor to the skin using hypoallergenic double-sided tape (3M 9917) was used to minimize cloth artifact and keep the sensor invisible.

Algorithm description

All data processing was performed using Matlab (Mathworks, Natick, MA, USA). The following steps were taken in the analysis process: preprocessing, removal of nonwear signals, semiautomatic calibration, and classification using a decision tree based on different signal features. During the preprocessing step the data was smoothed using a standard fourth order low-pass Butterworth filter with 5 Hz cut-off frequency. Since we were solely interested in identifying movements, and not in quantitative peak amplitudes, this smoothing does not affect the detection quality of the algorithm.

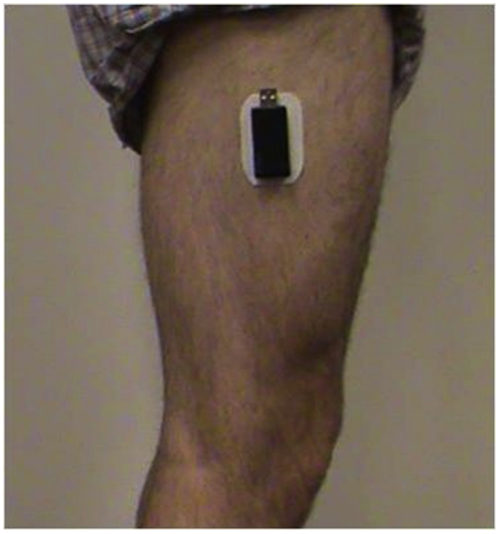

Figure 1 Attachment of device.

Calibration

After removing the obvious parts of the signal where the device was not worn, a calibration was performed by manually selecting a period of level walking in the subject's data set. The criterion for this period was at least five visible repetitions of the typical signal trace corresponding to one gait cycle. From this period of walking three parameters were automatically extracted: gait cycle frequency [GCF (Hz)], the average magnitude of the acceleration vector [T_1 (g)], and the average (low pass filtered) anteroposterior component of the signal [x-offset (g)]. This was the only calibration step in the process, making it user friendly. The calibration allows the general algorithm to adapt to variations in user height, morphology, sensor positioning, walking styles, and speed caused by limited range of joint motion, pain, or the use of walking aids, without the need to perform specific calibration movements in the laboratory.

Classification

The actual classification algorithm was a decision tree, with event-based windowing and decisions based on heuristic features [26]. At first, periods were divided into static and dynamic periods. Static periods were then classified as either resting or standing. Dynamic periods were divided into sedentary and upright parts, after which the sedentary periods were classified as either cycling or noncycling. Noncycling events were classified as resting, because no sedentary dynamic activities were considered in this study. In upright parts, a division was made between periods with a cyclic movement (locomotion) and periods without. The latter was classified as standing, the former was again divided into walking on a level surface and walking on stairs. Finally, a distinction was made between going upstairs and downstairs. This is schematically shown in Figure 2. The various decision steps are described in more detail below.

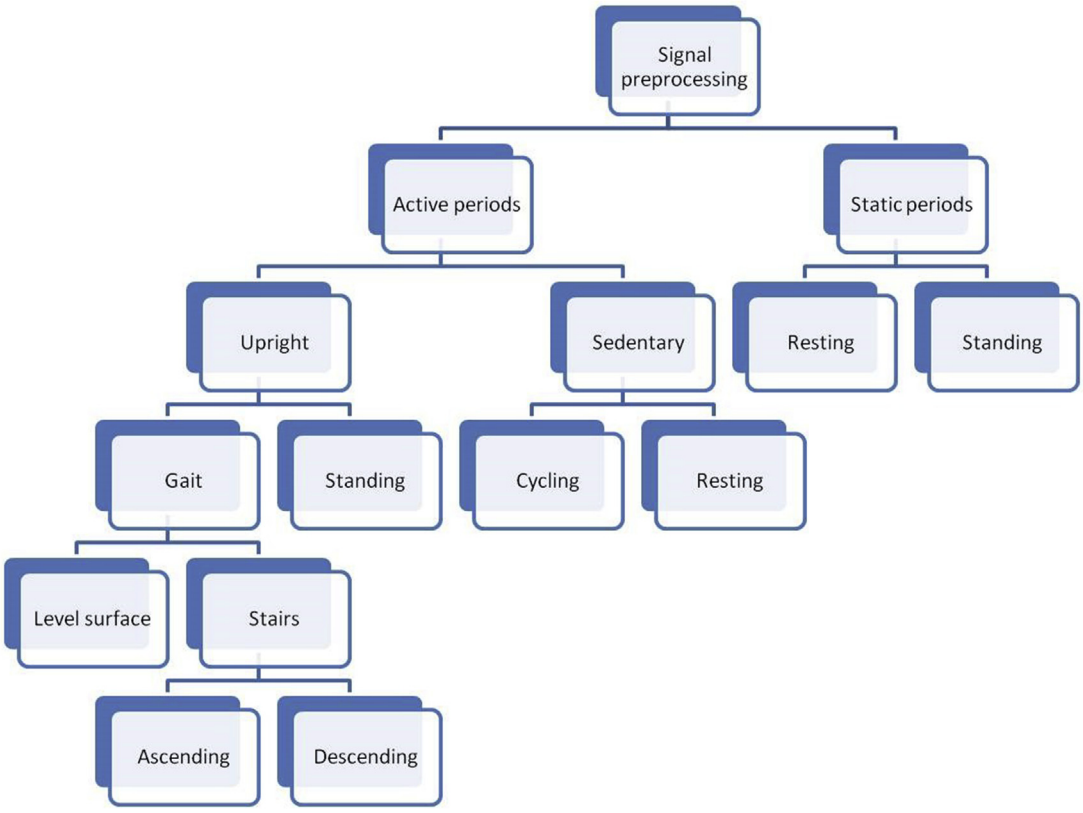

Figure 2 Decision tree used to classify different activities and postures.

Static or dynamic

The discrimination between static and dynamic periods was based on a 1-second average of the magnitude of the acceleration vector. This signal magnitude area (SMA) is calculated after high pass filtering.

$$SMA(i) = \int_{i-0.5}^{i+0.5} \sqrt{a_x^2(t) + a_y^2(t) + a_z^2(t)}\,dt, \quad (1)$$

where *SMA(i)* is the magnitude in 1-second window i, and a_x is the high pass filtered acceleration signal in the direction of the device's x-axis. A similar approach had already been used by others [25,27]. The threshold T_1, obtained from the calibration, was used to separate static (SMA $< T_1$) from dynamic (SMA $\geq T_1$) periods. Consecutive static and dynamic seconds were then grouped together to form either static or dynamic events. Two dynamic events interrupted by a static event shorter than 5 seconds were considered to be one dynamic event. This way each event was classified, instead of classifying every 1-second window separately.

Sitting or standing

Static events were classified as either standing or resting. This decision step was based on the inclination of the device. This approach has been used in a commercially available device emulated by Godfrey et al [15] Sit–stand transitions are now counted as changes in posture from sitting to standing or from sitting to an active upright event (usually locomotion) (Figure 3).

Sedentary or upright dynamic events

By applying a low pass filter (4^{th} order Butterworth, cut-off frequency 0.15 Hz) on the signal in dynamic events, the dynamic (inertial) characteristics in these events were removed and the static component (inclination) remained. Similar to static events (Figure 3), the dynamic events were divided into sedentary and upright dynamic events. The unfiltered signal was later used to classify these events further.

Locomotion

For upright dynamic events, an additional check was performed to distinguish locomotion from standing, shuffling, and other nonrhythmic upright activities (all classified as standing). Locomotion produces distinct patterns in both vertical and anteroposterior directions. The visibility of those patterns varies per subject, and thus the most visible one will be used for classification. In individuals who walk slowly, especially those using walking aids, the anteroposterior signal is more pronounced than the vertical acceleration signal. Consequently, for these individuals the event was classified as walking when there were at least five consecutive peaks in the anteroposterior signal interspaced less than 3 seconds (minimum cadence of 40 steps per minute), with a peak amplitude of 0.1g after smoothing (low-pass Butterworth filter with cut-off frequency of 2 Hz). Compton et al [28] defined slow walking as $<$ 80 steps/min. Therefore, we defined a GCF of $<$ 0.67 Hz as slow walking. For individuals who walk faster, a similar approach was used with different parameters. This time there had to be five consecutive peaks in the smoothed

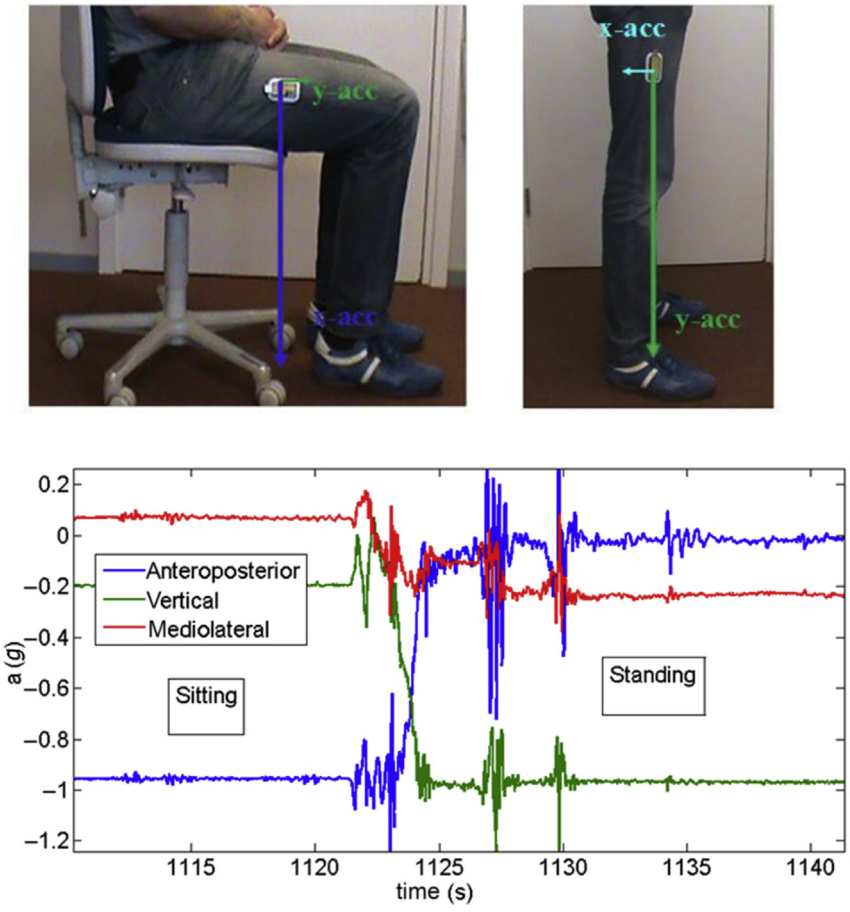

Figure 3 Inclinometer function of the device for posture detection.

vertical acceleration (low-pass Butterworth, cut-off frequency 4 Hz), with minimum amplitude of 0.2*g*. The minimum amplitudes were chosen in such a way that most peaks were found by Matlab's findpeaks function, while using the "MinPeakDistance" option in combination with the low-pass filter to limit the amount of double peaks. The consequence of allowing only periods with at least five consecutive peaks is that periods with only four steps or less will be classified as standing or shuffling. The number of steps in each walking bout is found by counting the number of peaks. In the vertical signal each peak corresponds to one step, in the anteroposterior signal each peak corresponds to a complete gait cycle, i.e., two steps. Cadence was consequently calculated as the inverse of the median interval (in minutes) between peaks.

Stairs

A characteristic signal feature which differentiates managing stairs and both stair ascent and descent from level walking or cycling was identified. Typically, when walking on stairs, individuals flex their knee and hip more than when walking on a level surface. For patients with a unilateral complaint (e.g., recovering from surgery) this effect is even more pronounced, because of the default sensor placement on the nonaffected leg. Those patients are often not able to walk stairs using a step-over-step strategy. They always lead with the nonaffected leg when ascending and the affected leg when descending, in order to keep the load on the affected leg to a minimum.

This meant that, averaged over one gait cycle, the axis parallel to the femur, and therefore the accelerometer, will be characteristically more inclined towards horizontal than during walking on a level surface. This so-called dynamic inclinometer feature can be best observed in the anteroposterior component of the signal after applying a low pass filter. To obtain the best temporal resolution the cut-off frequency should be as high as possible, but low enough to smoothen individual steps. Therefore, the cut-off frequency for GCF $>$ 0.67 Hz was 0.3 Hz, and for GCF $<$ 0.67 Hz the cut-off was 0.15 Hz. Assuming a minimum duration for a stair event, dynamic events longer than 4 seconds (GCF $>$ 0.67), or longer than 6 seconds (GCF $<$ 0.67), where the signal exceeded x-offset minus a threshold ($T_2 = 0.05g$ for GCF $>$ 0.67, and $T_2 = 0.1g$ for GCF $<$ 0.67), were classified as stair events (Figure 4).

Stair events were further classified as either ascending or descending by assessing the shape of the original signal. For GCF $>$ 0.67 Hz, this was done solely based on the vertical component of the signal, based on relative amplitude of neighbouring peaks. Empirically it was found that for ascending stairs a gait cycle consists of one higher and one lower peak, whereas for descending all peaks are near equal in amplitude (Figure 5). For GCF $<$ 0.67 Hz, both anteroposterior and vertical signals were used. It was found

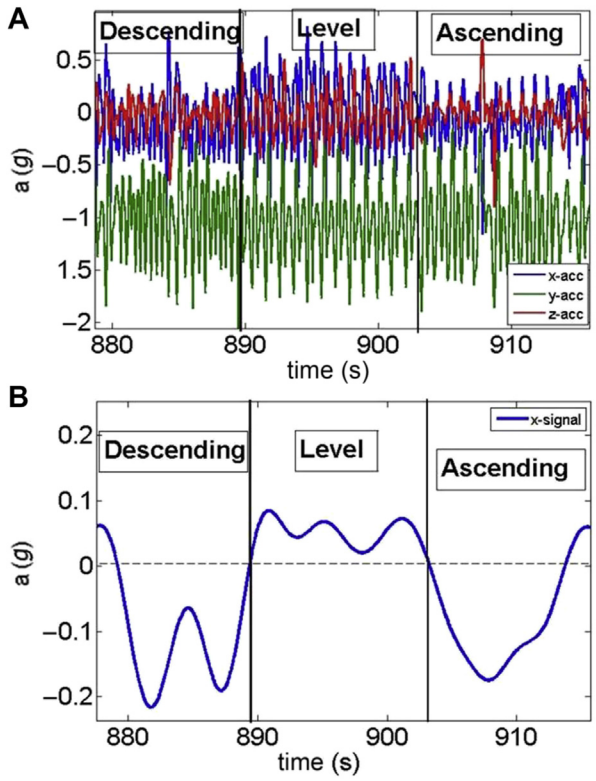

Figure 4 Typical (after smoothing) signal representing (A) stair events and level walking; and (B) low pass filtered anteroposterior signal of the same period. Dashed line indicates x-offset minus 0.05*g*.

that when ascending stairs the vertical signal peaks occurred right before an anteroposterior signal peak, whereas for descending the vertical peaks occurred right after anteroposterior signal peaks (Figure 5). These characteristic features were used in the algorithm to distinguish ascending from descending stairs.

Cycling

For sedentary dynamic events, only two possible activities were considered, cycling and noncycling, which was classified as resting. To classify a sedentary period as cycling there had to be at least five consecutive peaks in both the anteroposterior and the vertical acceleration, interspaced less than 1.5 seconds (corresponding to a cadence of 40 revolutions/min). The minimum amplitude of both sets of peaks was 0.1*g* after smoothing (low pass Butterworth, cut-off frequency 2.5 Hz). Cycling events were grouped together when pedalling was interrupted by a maximum of 20 seconds so that a cycling event will consist of parts with and without active pedalling.

Algorithm validation

Validation protocol

Participants ($n = 56$) were divided into two groups. The healthy group consisted of 16 individuals (mean age, 49 ± 20 years; mean body mass index, 23 ± 0.7 kg/m^2; male:female, 10:6). The patient group consisted of 40 individuals (mean age, 65 ± 9 years; mean body mass index, 30 ± 6 kg/m^2; male:female, 18:22) who underwent total joint arthroplasty 3—14 days prior to the test, thus representing individuals with limited functional capacity, in

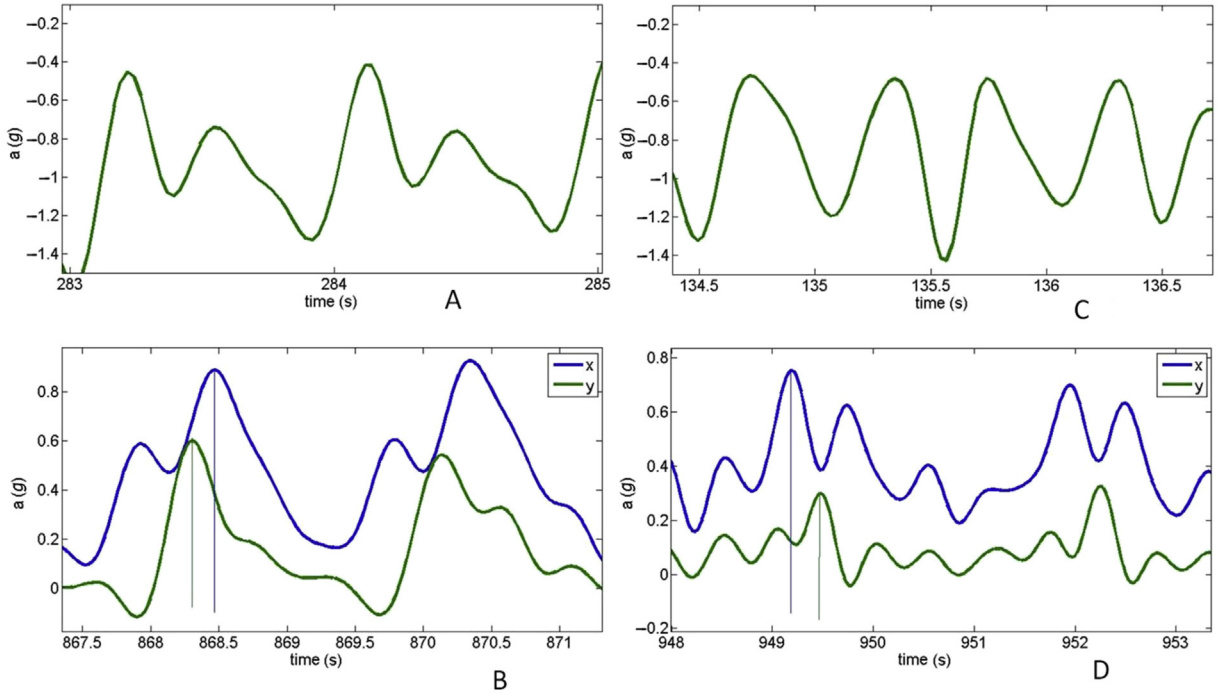

Figure 5 (A, B) Typical (after smoothing) stair ascending signal; and (C, D) descending signal. A and C are 'step-over-step' signals, B and D are 'step-by-step' signals.

order to challenge the robustness of the algorithm. All postoperative patients could mobilize safely, with some using walking aids such as crutches or a walker. More patients than healthy individuals were tested to reflect the higher movement variability in this group.

Healthy participants were followed for a period of 20–30 minutes, in which they could walk, sit, stand, go up and down the stairs, and use a bicycle to ride around. The order, pace, and duration of the activities were free and thus varied between individuals. Participants were free to roam about outdoors in a large space to create a realistic test environment.

This free test protocol (free order and duration of activities) is more realistic and demanding than common protocols, where activities are prescribed to the test participants in fixed order and for a long fixed minimum duration (e.g., 30 seconds)[24,29,30].

Postoperative patients performed as many activities (walking, getting up from a chair, and stair climbing) as they deemed possible. Typically, they climbed and descended the stairs, got up from a chair once or twice, and did several separate walking bouts. They walked and used the stairs using walking aids (crutches), under supervision of a physiotherapist. All activities were performed at self-selected pace and intensity.

All participants from both groups were recorded on video for post hoc identification of their movements, which served as the reference for the validation of the algorithms. Post hoc video analysis is used as the gold standard in other AM studies [31–33] as it allows the most accurate activity classification and timing, the possibility of reanalysis, or measurement of observer reliability.

Output parameters and accuracy

For the entire period the sensor was worn one of the following activities or postures was assigned by the video observer and algorithm: standing, resting, walking (level), ascending stairs, descending stairs, or cycling. In addition, the number of steps (for walking and stair events) was provided. The output parameters were event counts and durations for each activity. Observer and AM event counts were compared in a confusion matrix, comparable to Ermes et al [29]. Durations and number of steps were compared using Bland–Altman plots. In addition, the mean, standard deviation, and maximum and minimum error were calculated. The error per event count was calculated as:

$$E_x = \frac{|\#events_{AM} - \#events_{video}|}{\#events_{video}}, \quad (2)$$

where *#events* is the number of events of posture or activity *x*.

The error in durations was calculated as:

$$E_x = \frac{|duration_{AM} - duration_{video}|}{duration_{video}}, \quad (3)$$

where $duration_{AM}$ is the total duration of all correctly classified events *x*.

Results

Healthy group

The validation protocol for healthy individuals took on average 34 minutes per person. During this time the participants sat down, stood, walked, ascended and descended stairs, and rode a bicycle (Table 1).

A total of 992 different events (postures or activities) were identified by the video observer. A total of 986 of those 992 events were correctly classified by the algorithm. In six events the algorithm either partly or completely classified the event incorrectly. All standing and resting events were identified correctly, but one cycling and one walking event were (partly) classified as standing or resting. Two stair ascending events were misclassified (stair descent and level walking), and one stair descending event was misclassified (ascent; Table 2).

The mean error for the six event count categories varied between 0% (for cycling) and 2.8% for ascending stairs, corresponding to an accuracy of $>$ 97%. All sit–stand transitions were correctly identified. The mean relative error in step counting ranged from 1.7% for level walking to 6.4% for ascending stairs (Table 3). Figure 6 shows a Bland–Altman plot for step detection. Negative errors indicate an overestimation by the AM algorithm. The 95% limits of agreement were from −40 steps to 40 steps for level walking, from −14 steps to 20 steps for ascending stairs, and from −5 steps to 16 steps for descending stairs.

The agreement between AM and video for duration was excellent, with a mean error varying from 2.1% for resting to

Table 1 Mean, standard deviation, and range of activities performed by healthy individuals and patients.

Healthy	Time (min)					SST (*n*)	Bouts (*n*)			Steps (*n*)			Cadence (steps/min)
	Total	Rest	Walk	Stand	Cycle		Level	Ascent	Descent	Level	Ascent	Descent	
Mean	34	11	9	9	5	9	31	9	9	703	90	90	98
SD	8	4	2	3	1	2	7	2	2	151	21	20	8
Minimum	19	5	6	4	4	5	18	5	5	462	50	50	87
Maximum	43	16	12	13	7	10	37	11	10	910	110	100	112
Patients													
Mean	16	6	5	6	NA	3	4	1	1	108	10	10	56
SD	5	3	2	2	NA	1	1	0	0	46	1	1	10
Minimum	8	1	2	3	NA	1	2	1	1	57	9	9	40
Maximum	29	12	12	9	NA	4	7	2	1	267	11	11	74

NA = not applicable; SD = standard deviation; SST = sit-stand transitions.

Table 2 Confusion matrix, total number of correctly and incorrectly classified events.

Healthy								Patient						
↓ True \ AM→	Stand	Rest	Walk level	Stairs ascent	Stairs descent	Cycling	Total	↓ True \ AM→	Stand	Rest	Walk level	Stairs ascent	Stairs descent	Total
Stand	86	0	0	0	0	0	86	Stand	58	0	0	0	0	58
Rest	0	69	0	0	0	0	69	Rest	0	102	0	0	0	102
Walk Level	1	0	496	0	0	0	497	Walk level	2	0	152	0	0	150
Stairs ascent	0	0	1	144	1	0	146	Walk ascent	2	0	0	42	0	40
Stairs descent	0	0	0	1	145	0	146	Walk descent	0	0	0	0	40	40
Cycling	1	1	0	0	0	48	48							

AM = activity monitoring.

Table 3 Mean, standard deviation (SD) and range of counting error in output parameters.

Healthy										
Activity	Stand	Rest	Level Walk	Stairs ascent	Stairs descent	Cycling	SST	Steps level	Steps ascent	Steps descent
Mean error (%)	1	1.3	0.2	2.8	2.4	0	0	1.7	6.4	5.4
SD	4.2	5	0.8	7.9	4.9	0	0	1.5	5.6	5.5
Minimum	0	0	0	0	0	0	0	0	4.1	1.1
Maximum	16.7	20	3.2	20	11	0	0	4.4	16.6	17.4
Patients										
Activity	Stand	Rest	Level walk	Stairs ascent	Stairs descent	Cycling	SST	Steps level	Steps ascent	Steps descent
Mean error (%)	7.5	0.6	1.2	3.3	0	NA	0.6	3.4	6.9	8.2
SD	24.2	4	5.7	14.7	0	NA	4	3.6	18.4	15.4
Minimum	0	0	0	0	0	NA	0	0	0	0
Maximum	100	25	33.3	100	0	NA	25	16.7	66.7	66.7

NA = not applicable; SD = standard deviation; SST = sit-stand transitions.

2.8% for cycling. Figure 7 shows Bland–Altman plots with the error in duration. On average, walking duration was slightly underestimated, while resting duration was overestimated.

Patient group

For patients the mean time spent performing the validation protocol was 16 ± 5 minutes. All patients performed at least two walking bouts and ascended and descended the stairs once. Cycling was not part of the mobilization routine prescribed by the physiotherapist, and was thus not performed in this group (Table 1).

A confusion matrix was constructed in the same way as for healthy participants. A total of 390 events were identified by the video observer. In four events (2 level walking and 2 stair ascending) a pause (classified as standing) was detected by the algorithm, leading to four erroneous standing events. All other events were correctly classified (Table 2).

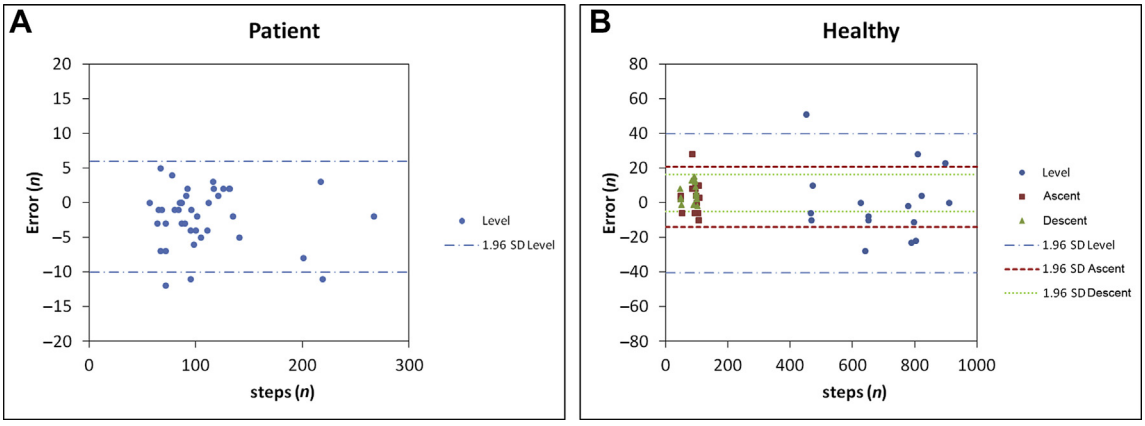

Figure 6 Bland–Altman plots showing the error in step detection and 95% limits of agreement (1.96 SD) for (A) patients; and (B) healthy individuals. SD = standard deviation.

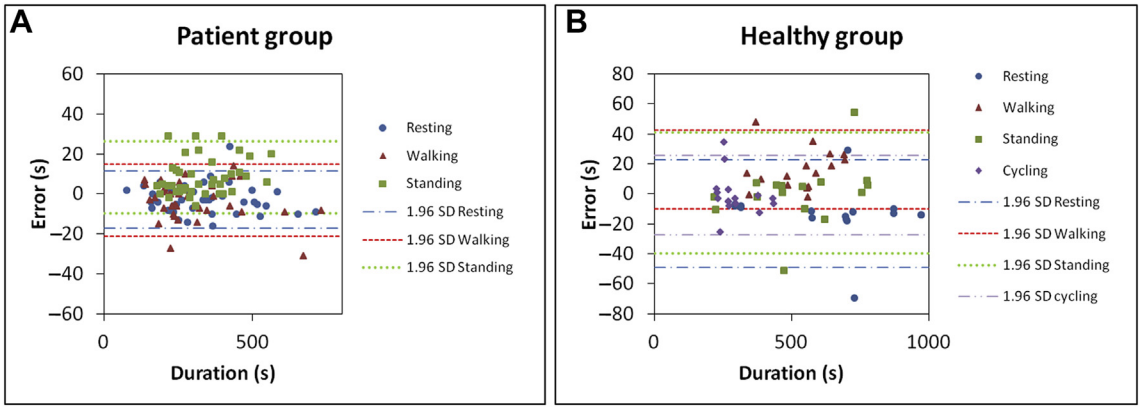

Figure 7 Bland–Altman plots showing the error in event duration and 95% limits of agreement (1.96 SD) for (A) patients; and (B) healthy individuals. SD = standard deviation.

The mean error regarding event counts in patients varied between 0% (for descending stairs) and 7.5% for standing, corresponding to an accuracy of > 92%. Only one sit–stand transition was missed by the algorithm (mean error 0.6%). The mean relative error in steps ranged from 3.4% for level walking to 8.2% for descending stairs (Table 3). Figure 6 shows the Bland–Altman plot for step detection. Only level steps are shown because of the small number of steps in stair events per patient. The 95% limits of agreement were −10 steps and +5 steps.

The mean error in duration varied from 2.0% for resting to 2.9% for standing. Figure 7 shows the error in duration. On average, standing duration was underestimated, whereas walking and resting duration were overestimated.

Discussion

This study presented a combination of previously described and new AM classifiers, such as the dual-axis approach for gait and the averaged hip flexion for stairs. A semi-automated calibration method was used to allow the sensor and algorithm to be used in healthy individuals with a wide range of functional capacity or limitations from direct postoperative arthroplasty, including the use of walking aids.

This novel combination of AM algorithms was validated in a demanding and realistic test protocol allowing free choice of activity type, order, and duration.

With an average detection accuracy ranging between 92% and 100% for classifying all activity events and participants in a demanding laboratory validation set-up, it was shown that the method meets and exceeds values commonly accepted for free field use of AM and commercial devices used in clinical studies [24,27,29,31–36].

The comparison of accuracy values in other studies is challenged by the fact that different validation protocols were used as there is no standard or consensus about it. Accuracy values are influenced by the choice and definition of activities to be identified, the way the test protocol is set up, the participants used in the study, and the calculation of accuracy values themselves.

Nyan et al [37] and Muscillo et al [38] presented algorithms to classify locomotion (level, ascending stairs, descending stairs) and reported > 95% accuracies, comparable to the accuracy found in this study (97%). Fortune et al [31] reported 92% accuracy in step counts for walking, stair events, and jogging events. In their study the maximum average error was 8.2% in step counts for patients descending stairs. Early research by Mathie et al [30] reported accuracy values of 94–98% for resting/activity using a fixed protocol and counting a correct detection even when there was an overlap in time between classified event and actual event. Few groups have tried to identify a complete set of postures and activities as combined in this study. Ermes et al [29] reported values between 78% (Nordic walking) and 99% (resting) in a confusion matrix comparable to this study, however they used multiple sensors. Khan et al [24] used a single sensor, independent of sensor positioning, and reported 94–98% accuracy for the same activities as in this study, except sit–stand transitions. Similar results were achieved in this study, using a single fixed sensor position. A position-independent algorithm would be advantageous over a fixed position, but its validity has yet to be proven in a simulated free-living environment such as in the protocol applied in this study. In the study by Khan et al [24], start and end points were manually annotated in the signal, whereas in a free-living environment transitions between activities pose a challenge. Ermes et al [29] reported that in an unsupervised free environment overall accuracy drops by up to 17%, and individual classification accuracies by up to 40%. Therefore, we propose that a validation protocol should resemble a real life situation as closely as possible. The Bland–Altman plots in Figures 6 and 7 show no correlation between absolute error and event duration or number of steps. In the validation protocol events were relatively short compared to daily life, which means that the relative error in a longer measurement will probably be even smaller than described in this study.

The studies mentioned above have all validated their algorithms using young, healthy individuals. Only Khan et al [24] tested on elderly, healthy individuals, which was their target population. However, functional limitations and neurological disorders affect movement. This challenges AM algorithms, and therefore the validation will be different [21–23]. Raymond et al [20] validated the PAL2 monitor (Gorman ProMed Pty Ltd, Victoria, Australia) for

different postures, transitions, and walking in an elderly population but did not include stair events. O'Donoghue and Kennedy [18] found high agreement using the activPal (Pal Technologies, Glasgow, UK) sensor versus video observation for sit—stand transitions and walking events, but low agreement for step counts. In this study, high accuracies were found for all categories including step counts and for both healthy and functionally affected groups, indicating a high robustness of the algorithm.

Accuracy figures were of equal quality in both the healthy group and the patient group, showing that the algorithm works well in a wide spectrum of functional capabilities. The good results in both groups show that the calibration and dual-axis approach perform well in practice, even when locomotion is performed at different speeds or styles (e.g., walking aids, step-over, or step-by-step stair climbing). It does, however, mean that the investigator has to provide some manual input to the algorithm when analysing the data set, by selecting one walking event in a simple and fast step taking less than 1 minute.). This is a practical disadvantage to the algorithms where a completely automated calibration is performed [24,31].

The highest accuracy in both groups was found in sit—stand transitions, which can be attributed to the robust inclinometer function of the device, and placement on the upper leg.

The lowest accuracy (5—8% error) was found for step detection on stairs. This can be explained by the fact that the staircases in the validation protocol were short (10 stairs vs. 15—20 for a normal flight). One step before and after the stair event can easily be counted as a step on stairs, while the first or last step can be counted as a level step. We accept this error, because if and how many times an individual can manage stairs is more relevant to a clinician than the exact amount of steps on the staircase.

Instead of reporting accuracy in event counts, is it then better to report timing accuracy of activity classifications, as reported by Nyan et al [37] In clinical practice, both accuracy figures can be relevant and have been calculated in this study. Timing error or accuracy seems more relevant for activities or postures that take up a large portion of the day, such as resting, standing, walking, or cycling, and event based errors seem more meaningful for short events such as sit—stand transition, managing stairs, and individual steps.

In this study, not only the mean errors for all participants, but also means and ranges for each individual are reported. Especially in the patient group, the maximum error per individual could be large (up to 100%). However, patients could only perform a small amount of activity events. One misclassified event can already lead to large relative errors, which do not reflect the true accuracy if more activity bouts could have been tested per individual. This validation protocol was designed to collect a large number of events overall from a large number and variety of individuals.

The major limitation of almost all validation studies is the translation to free-living conditions. Ermes et al [29] reported that the overall accuracy of their algorithm dropped by 17% overall when participants were left unsupervised. This emphasizes the need for a validation protocol that mimics real life as close as possible. Fortune et al [31] asked their participants to fidget while sitting or standing to challenge their algorithms. In this study, we chose not to, considering it too unnatural. Ideally participants would be followed and video-recorded for a longer period than in this test protocol (e.g., a day). This, however, would lead to practical and ethical difficulties, as video observation would be too time consuming and invasive, while self-report is too tedious and inaccurate.

Conclusion

It was shown that with a single, lightweight, three-dimensional accelerometer fixed to the body and a computationally simple postprocessed signal analysis it is possible to design an AM tool satisfying three important criteria. Firstly, it is able to detect and differentiate a large set of daily life activities relevant to orthopaedic patient assessment with high accuracy. Secondly, it performs equally well for healthy individuals as well as motorically limited patients because of the semiautomatic calibration step. Finally, it is validated using a challenging validation protocol.

The algorithms are independent of a particular sensor and will work on any three-dimensional accelerometer with a measurement range and sampling frequency similar to the device used in this study. The low-cost sensor and simple methods presented in this study can thus be applied, for instance, to establish reference databases of habitual activity levels of specific patient groups, to objectify outcome assessment following orthopaedic intervention, to monitor compliance to therapies involving activity, individualize recovery programs, or to power scientific studies about treatment alternatives.

Conflicts of interest

The authors declare that they have no conflict of interest. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Ni Scanaill C, Carew S, Barralon P, Noury N, Lyons D, Lyons GM. A review of approaches to mobility telemonitoring of the elderly in their living environment. Ann Biomed Eng 2006;34:547—63.

[2] Celler BG, Lovell NH, Hesketh T, Ilsar ED, Earnshaw W, Betbeder-Matibet L. Remote home monitoring of health status of the elderly. Medinfo 1995;8:615—9.

[3] Terwee CB, Bouwmeester W, van Elsland SL, de Vet HC, Dekker J. Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. Osteoarthritis Cartilage 2011;19:620—33.

[4] Dobson F, Hinman RS, Roos EM, Abbott JH, Stratford P, Davis AM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthritis Cartilage 2013;21:1042—52.

[5] Pitta F, Troosters T, Spruit MA, Probst VS, Decramer M, Gosselink R. Characteristics of physical activities in daily life in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2005;171:972—7.

[6] Plasqui G, Westerterp KR. Physical activity assessment with accelerometers: an evaluation against doubly labeled water. Obesity 2007;15:2371—9.
[7] Westerterp KR. Assessment of physical activity: a critical appraisal. Eur J Appl Physiol 2009;105:823—8.
[8] Grimm B, Bolink S. Evaluating physical function and activity in the elderly patient using wearable motion sensors. EFORT Open Reviews 2016;1:112—20.
[9] Protopapadaki A, Drechsler WI, Cramp MC, Coutts FJ, Scott OM. Hip, knee, ankle kinematics and kinetics during stair ascent and descent in healthy young individuals. Clin Biomech (Bristol, Avon) 2007;22:203—10.
[10] Oiestad BE, Osteras N, Frobell R, Grotle M, Brogger H, Risberg MA. Efficacy of strength and aerobic exercise on patient-reported outcomes and structural changes in patients with knee osteoarthritis: study protocol for a randomized controlled trial. BMC Musculoskelet Disord 2013;14:266.
[11] Storti KL, Pettee KK, Brach JS, Talkowski JB, Richardson CR, Kriska AM. Gait speed and step-count monitor accuracy in community-dwelling older adults. Med Sci Sports Exerc 2008; 40:59—64.
[12] Aminian K, Robert P, Buchser EE, Rutschmann B, Hayoz D, Depairon M. Physical activity monitoring based on accelerometry: validation and comparison with video observation. Med Biol Eng Comput 1999;37:304—8.
[13] Annegarn J, Spruit MA, Uszko-Lencer NH, Vanbelle S, Savelberg HH, Schols AM, et al. Objective physical activity assessment in patients with chronic organ failure: a validation study of a new single-unit activity monitor. Arch Phys Med Rehabil 2011;92. 1852—7 e1.
[14] de Groot S, Nieuwenhuizen MG. Validity and reliability of measuring activities, movement intensity and energy expenditure with the DynaPort MoveMonitor. Med Eng Phys 2013;35: 1499—505.
[15] Godfrey A, Culhane KM, Lyons GM. Comparison of the performance of the activPAL professional physical activity logger to a discrete accelerometer-based activity monitor. Med Eng Phys 2007;29:930—4.
[16] Skotte J, Korshoj M, Kristiansen J, Hanisch C, Holtermann A. Detection of physical activity types using triaxial accelerometers. J Phys Act Health 2014;11:76—84.
[17] Mukhopadhyay SC. Wearable sensors for human activity monitoring: a review. IEEE Sens J 2015;15:1321—30.
[18] O'Donoghue D, Kennedy N. Validity of an activity monitor in young people with cerebral palsy gross motor function classification system level I. Physiol Meas 2014;35:2307—18.
[19] Laudanski A, Brouwer B, Li Q. Activity classification in persons with stroke based on frequency features. Med Eng Phys 2015; 37:180—6.
[20] Raymond M, Winter A, Holland AE. Validation of an activity monitor in older inpatients undergoing slow stream rehabilitation. J Phys Act Health 2015;12:1298—303.
[21] Brandes M, Ringling M, Winter C, Hillmann A, Rosenbaum D. Changes in physical activity and health-related quality of life during the first year after total knee arthroplasty. Arthritis Care Res 2011;63:328—34.
[22] Cyarto EV, Myers A, Tudor-Locke C. Pedometer accuracy in nursing home and community-dwelling older adults. Med Sci Sports Exerc 2004;36:205—9.
[23] Langer D, Gosselink R, Sena R, Burtin C, Decramer M, Troosters T. Validation of two activity monitors in patients with COPD. Thorax 2009;64:641—2.
[24] Khan AM, Lee YK, Lee S, Kim TS. Accelerometer's position independent physical activity recognition system for long-term activity monitoring in the elderly. Med Biol Eng Comput 2010;48:1271—9.
[25] Godfrey A, Conway R, Meagher D, OLaighin G. Direct measurement of human movement by accelerometry. Med Eng Phys 2008;30:1364—86.
[26] Preece SJ, Goulermas JY, Kenney LP, Howard D, Meijer K, Crompton R. Activity identification using body-mounted sensors—a review of classification techniques. Physiol Meas 2009;30:R1—33.
[27] Mathie MJ, Coster AC, Lovell NH, Celler BG. Detection of daily physical activities using a triaxial accelerometer. Med Biol Eng Comput 2003;41:296—301.
[28] Compton RO, Ulcak M, Gonzales JU. The acute effect of fast and slow stepping cadence on regional vascular function. Int J Sports Med 2015;36:1041—5.
[29] Ermes M, Parkka J, Mantyjarvi J, Korhonen I. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. IEEE Trans Inf Technol Biomed 2008;12:20—6.
[30] Mathie MJ, Celler BG, Lovell NH, Coster AC. Classification of basic daily movements using a triaxial accelerometer. Med Biol Eng Comput 2004;42:679—87.
[31] Fortune E, Lugade V, Morrow M, Kaufman K. Validity of using tri-axial accelerometers to measure human movement - Part II: Step counts at a wide range of gait velocities. Med Eng Phys 2014;36:659—69.
[32] Lugade V, Fortune E, Morrow M, Kaufman K. Validity of using tri-axial accelerometers to measure human movement - Part I: Posture and movement detection. Med Eng Phys 2014;36: 169—76.
[33] De Vries SI, Garre FG, Engbers LH, Hildebrandt VH, Van Buuren S. Evaluation of neural networks to identify types of activity using accelerometers. Med Sci Sports Exerc 2011;43: 101—7.
[34] Parkka J, Ermes M, Korpipaa P, Mantyjarvi J, Peltola J, Korhonen I. Activity classification using realistic data from wearable sensors. IEEE Trans Inf Technol Biomed 2006;10: 119—28.
[35] Duncan S, White K, Mavoa S, Stewart T, Hinckson E, Schofield G. Active transport, physical activity, and distance between home and school in children and adolescents. J Phys Act Health 2016;13:447—53.
[36] Esbensen BA, Thomsen T, Hetland ML, Beyer N, Midtgaard J, Loppenthin K, et al. The efficacy of motivational counseling and SMS-reminders on daily sitting time in patients with rheumatoid arthritis: protocol for a randomized controlled trial. Trials 2015;16:23.
[37] Nyan MN, Tay FE, Seah KH, Sitoh YY. Classification of gait patterns in the time-frequency domain. J Biomech 2006;39: 2647—56.
[38] Muscillo R, Schmid M, Conforto S, D'Alessio T. An adaptive Kalman-based Bayes estimation technique to classify locomotor activities in young and elderly adults through accelerometers. Med Eng Phys 2010;32:849—59.