

Med Sci Sports Exerc. Author manuscript; available in PMC 2014 November 01.

Published in final edited form as:

Med Sci Sports Exerc. 2013 November; 45(11): 2193-2203. doi:10.1249/MSS.0b013e31829736d6.

Activity recognition using a single accelerometer placed at the wrist or ankle

Andrea Mannini¹, Stephen S. Intille², Mary Rosenberger³, Angelo M. Sabatini¹, and William Haskell³

¹The BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa, Italy

²College of Computer and Information Science and Bouvé College of Health Sciences, Northeastern University, Boston, MA, United States

³Stanford Prevention Research Center, Stanford University, Stanford, CA, United States

Abstract

PURPOSE—Large physical activity surveillance projects such as the UK Biobank and NHANES are using wrist-worn accelerometer-based activity monitors that collect raw data. The goal is to increase wear time by asking subjects to wear the monitors on the wrist instead of the hip, and then to use information in the raw signal to improve activity type and intensity estimation. The purpose of this work is obtaining an algorithm to process wrist and ankle raw data and classify behavior into four broad activity classes: ambulation, cycling, sedentary and other.

METHODS—Participants (N = 33) wearing accelerometers on the wrist and ankle performed 26 daily activities. The accelerometer data were collected, cleaned, and preprocessed to extract features that characterize 2 s, 4 s, and 12.8 s data windows. Feature vectors encoding information about frequency and intensity of motion extracted from analysis of the raw signal were used with a support vector machine classifier to identify a subject's activity. Results were compared with categories classified by a human observer. Algorithms were validated using a leave-one-subject-out strategy. The computational complexity of each processing step was also evaluated.

RESULTS—With 12.8 s windows, the proposed strategy showed high classification accuracies for ankle data (95.0%) that decreased to 84.7% for wrist data. Shorter (4 s) windows only minimally decreased performances of the algorithm on the wrist to 84.2%.

CONCLUSIONS—A classification algorithm using 13 features shows good classification into the four classes given the complexity of the activities in the original dataset. The algorithm is computationally-efficient and could be implemented in real-time on mobile devices with only 4 s latency.

Keywords

Activity classification; inertial sensing; mobile health; leave-one-subject-out validation; activity measurement; energy expenditure

Copyright © 2013 American College of Sports Medicine

Corresponding Author. Andrea Mannini, Polo Sant'Anna Valdera, viale Rinaldo Piaggio 34, 56025 Pontedera (Pisa), Italy Tel: +39-050-88-3410, fax: +39-050-88-3101, a.mannini@sssup.it.

Conflicts of Interest. The authors have no conflicts of interest to disclose.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Accurate quantification of daily physical activity and energy expenditure would advance science and assist with proper management of pathologies such as obesity, diabetes and cardiovascular diseases (21). Accelerometer-based activity monitors are capable of quantifying human motion, covering the range of acceleration amplitudes and frequencies required to measure human movement (4). Moreover, their low power consumption, small dimensions, and light weight contribute to ease of wearability and make long term activity monitoring practical. Accelerometer-based activity monitors often integrate raw accelerometer data over short windows of time ("epochs") into "activity counts" (28). Activity counts summarize data in an epoch, simplifying data management, analysis and interpretation; however, information about the structure of the raw accelerometer signal is lost that might be used by algorithms to infer specific activity type (3, 17), gait parameters (gait phase detection, walking speed estimation) (16, 23), posture transition and balance (18, 33), rehabilitation progress (20) and the detection of falls (5, 9). Activity type information inferred from raw data analysis can be used to improve estimates of energy expenditure (1, 10) or walk/run speed (26) by switching regression parameters to those tuned to the recognized activity.

Activity classification using accelerometers can be obtained using one or more sensors on the body. A multi-sensor configuration is preferable because it allows detecting a larger variety and finer complexity of activities by capturing both upper and lower body motion independently (3). Although multi-sensor systems are becoming more practical, singlesensor systems may simplify study administration and lower study administrative cost and are therefore current practice in larger studies. With a single sensor, common location choices are the hip, thigh, upper arm, wrist, or ankle. The hip location has been used extensively in physical activity measurement studies. It generally captures major body motions, but algorithms using hip data can underestimate overall expenditure on activities such as bicycling or arm ergometry, where the hip movement is not proportional to movement of the limbs (30). Hip sensors must be attached to the outside of clothing or worn on a belt and that can lead to lower wear-time compliance, especially during sleep. Water resistant wrist-worn devices can be used without discomfort during activities of daily living, including sleep, and can remain on even if clothes are changed and do not require a special belt or clip, thus leading to improved wear time (12). However, given the large amount of gesticulation involving upper limbs that does not correspond to large body movement (and therefore large energy expenditure), estimating energy expenditure from wrist movement counts may introduce more error than the same calculations at the hip (22). Nevertheless, the NHANES and UK Biobank surveillance studies switched from the hip to the wrist location in the most recent round of data collection to improve wear time (6); the intent is to log and use raw data analysis to characterize activity despite the additional wrist gesticulation. Preliminary data from the 2011-2014 NHANES study suggests that wrist placement will result in higher wear time (27), but processing the raw data to characterize overall activity provides a challenge because of the wrist gesticulation and variability in movement. For example, one study that compared different sensor placement sites reported lower activity type detection performance for the wrist with respect to other locations for the detection of sit-to-stand or lie-to-stand transitions (2). The wrist may move differently during the same activity depending on what is in the hand and what the hand is holding or stabilizing. For example, the raw motion signature for the wrist while walking holding a heavy bag, walking with a full cup, or walking with a mobile phone held to the head may differ. Several recent studies explored the problem of detecting activities from wrist-worn sensors (8, 14, 25, 31, 32). As described in the discussion section, the prior studies have limitations due to the

amount of data tested, the complexity of the activities tested, or the validation approach used when reporting results.

An alternative placement location that deserves consideration is just above the ankle, worn under clothing. Drawbacks cited for the ankle location are concerns about tight fitting boots and concerns that subjects will refuse to wear an ankle-worn device that may resemble an alcohol monitor or police tether. A sufficiently small and thin sensor may minimize these problems, however. As with the wrist location, a sensor could be attached to the ankle under the clothing and left on day and night, perhaps increasing wear time and compliance. Further, the ankle placement site may be particularly useful for pedestrian navigation and gait analysis purposes (24), such as gait segmentation (19) or walking speed estimation (15). Despite the potential benefits, we are unaware of prior work studying the impact of ankle placement on data analysis, wear time and acceptability.

In this work, we replicated a recently reported algorithm for detecting activity class from raw accelerometer data collected at the wrist (Zhang et al. (32)) and tested it on a dataset with 33 participants performing a set of daily physical activities. We aimed at classifying activities within four classes (ambulation, cycling, sedentary activities and other activities), using a leave-one-subject-out validation to characterize algorithm performance. Various combinations of window lengths (i.e., the amount of data acquired to give a single classification output) and feature sets (sets of variables used for classification purposes) were tested to develop an algorithm. The algorithm is computationally efficient and suitable for implementation on mobile phones that support real-time mobile phone interventions using wrist-mounted wireless accelerometers. Real time capabilities allow fast feedback to the user that may be useful in future systems designed to use this information to improve compliance.

Materials and methods

1. Data acquisition and annotation

This study uses a dataset of acceleration data tagged with activity type that was previously acquired for other studies from 42 adults recruited from the Stanford, CA community. Stanford University's Human Use Committee approved the protocol, and written informed consent was obtained from all subjects before participation. Triaxial accelerometers called Wockets (12) were attached using custom Velcro bands to each participant's ankle, thigh, hip upper arm and wrist. The wrist Wocket was placed on the dorsal aspect of the dominant wrist midway between the radial and ulnar process. The ankle Wocket was placed on the outside of the ankle, just above the lateral malleolus. The thigh Wocket was located on the anterior thigh midway between the top of the patella and the inguinal fold. The arm sensor was worn over the lateral side of the arm midway between the shoulder and the elbow. Arm and thigh Wockets were attached with both adhesive tape and a sleeve worn over the Wocket and around the sensor. The hip Wocket was worn on a belt around the participant's waist on the dominant side of the body. The Wockets are small, thin, and lightweight devices ($43 \text{mm} \times 30 \text{ mm} \times 7 \text{ mm}$, 13g) that include a triaxial accelerometer (MMA7331LCR1), a microprocessor (ATMEGA1284P), a Bluetooth radio (RN-41) and a rechargeable Lithium battery (3.7V 240mAh). They are optimized for long-term wearability for physical activity monitoring studies, where mobile phones are used for data collection. Raw acceleration data (range ± 4 g, g = 9.81 m/s²) were acquired at 90 Hz and sent using the Bluetooth wireless protocol to a smartphone.

The experimental protocol consisted of asking participants to perform a guided sequence of laboratory-based physical activities and simulated daily activities. Activities were annotated during the execution of tasks using a voice recorder, and then timings on the voice recording

were used to annotate start/stop times for specific activities being observed. Data and annotation were synchronized using custom software (12). Twenty-six activities with more than 0.5 min of steady state data were labeled in the original dataset. Those activities have been clustered into four more general categories for this study: sedentary (lying, sitting, internet search, reading, typing, writing, sorting files on paperwork, standing still), cycling (indoor and outdoor), ambulation (natural walking, treadmill walking, carrying a box, stairs up/down) and other activities (sweeping with broom, painting with roller or brush). In the current dataset other sedentary activities such as driving a car or riding public transit were not available. Data that were not labeled or for which the label was "unknown" were discarded. Multi-tasking behaviors were not allowed during experiments, except for the activity walking-carrying-a-load.

Data from 9 participants were discarded due to high data loss or to technical problems affecting the wrist or ankle sensor, as reported in notes taken by the staff at the time of data collection. Ankle and wrist data from the remaining 33 participants (11 males, 22 females, ages 18-75, height 168.5 ± 9.3 cm (range 149-189), weight 70.0 ± 15.6 kg (range 48-114)) were imported into the Mathworks Matlab (v7.6, Natick MA) environment, which was used for all evaluations described. All other available data were discarded as they were not pertinent to the aims of this study. The dataset and Matlab code used in this study are available to interested researchers [http://mhealth.ccs.neu.edu/datasets]. The dataset was acquired with a protocol designed to encourage natural behavior within a lab setting. Participants were told what to do but not how to do it, and staff annotated the activities as previously described.

This data collection procedure allows for natural participant variability in how activities are performed but can also lead to errors in annotation at activity transitions due to reaction time when labeling. For this reason, in this work we discarded one window (12.8 s) before and after each label transition. When smaller windows were considered, to keep the analysis consistent, 12 s before and after each transition were still discarded, corresponding to 3 or 6 windows for 4 s and 2 s windows respectively. Another type of error is that some short activity changes are not labeled at all. For example, the dataset contains examples where a participant stops briefly during non-treadmill walking, such as at a door that had to be opened. In such cases, even though the participant is standing still briefly, the label for the data is still "ambulation." Some errors can be detected by using the ankle acceleration recordings because in data labeled as "ambulation," the Signal Magnitude Vector (SMV),

 $SMV = \sqrt[2]{acc_x^2 + acc_y^2 + acc_z^2}$, of the ankle sensor must show movement. We therefore use the ankle sensor to identify these errors in labeling and correct labels indicating ambulation. In particular, 2 s windows labeled as "ambulation" with a SMV standard deviation less than 0.1 g are marked as labeling errors and discarded. In the results section, data with this correction are referred as "corrected data".

Data-loss due to Bluetooth wireless transmission errors was handled by discarding windows with less than the 80% of the number of expected samples at the nominal 90 Hz sampling rate. In such cases, a new window was started at the end of the data gap. Some fluctuations in the sampling rate may occur in the remaining windows due to the wireless connection. Before extracting frequency domain features, SMVs in each window were linearly interpolated to obtain the same number of samples in every window.

2. Data preprocessing and features evaluation

Three-axis raw accelerometer data were preprocessed to extract the SMVs according to the previously introduced definition. The resulting 90 Hz SMV signal is independent of the orientation of the sensing node. SMVs were low pass filtered using a 15 Hz cut-off 4th order

Butterworth filter to limit the bandwidth of the signal to the frequencies common in human motion (4), removing high frequency noise.

To classify data within the four defined activity classes, the SMVs data were broken into 12.8 s non overlapping windows. This window size was proposed by Zhang et al. (32). Window lengths of 4 s and 2 s were also tested, because shorter windows reduce latency in providing feedback in real-time implementations and would therefore be preferable. Windowed portions of signals were processed to extract features commonly used in raw data processing of accelerometer signals (3) and the feature set proposed by Zhang et al. (32). Mean and standard deviation of SMV were considered jointly with a time-frequency analysis of 12.8 s windows. The analysis of power spectral density is aimed at characterizing (a) total power in the frequencies between 0.3Hz and 15Hz, (b) first and second dominant frequencies and their powers in the same frequency band, (c) dominant frequency in the band 0.6-2.5 Hz and its power, (d) the ratio between the power of the first dominant frequency and the total power (0.3-15Hz), and (e) the ratio between the dominant frequency of the current segment and the segment before. We also considered two features based on wavelet transforms found to improve classification accuracy in Zhang et al. (32):

$$\begin{array}{l} -DWT_{\scriptscriptstyle SMV} \! = \! \! \sum_{j=\alpha}^{\beta} \! d_j^2/SMV^2 \\ -DWT_{\scriptscriptstyle SMV1} \! = \! \! \sum_{j=\alpha}^{\beta} \! d_j^2/\sum_{j=\alpha}^{J} \! d_j^2 \end{array}$$

Where d_j is the decomposed signal at level j of the SMV and $d_j^2 = d_j d_j^T$. The selected wavelet was the Daubechies "db10" for its close match to walking data (32). The maximum level considered for decomposition was J = 8, whereas the levels considered for the evaluation of these features were = 5 and = 6. We also introduced two simple features evaluated on each windowed portion of SMVs – the minimum and the maximum values within the window.

Computationally-simple features are preferred to those that require substantial processing such as wavelet analysis if the complex features only modestly improve results. Simple features minimize latencies and maximize battery life of computing devices that run the classification algorithms. Algorithms that permit real-time implementations may ultimately be used in measurement and intervention studies that provide real-time feedback to participants based on detected activity. To identify the best trade-off between accuracy and classifier complexity, eight different feature set combinations were evaluated. In particular, the effect of removing the wavelet based features from the training set was investigated.

3. Classifier validation

Preliminary testing showed that the highest classification rates were achieved using support vector machine (SVM) classifiers (29). The best outcomes were achieved using a *radial basis function* kernel with upper complexity bound C = 100 and = 0.1. SVM classifiers are desirable because the optimization criteria are convex, which implies that a global optimal solution exists (29), and many toolboxes exist that simplify application of the algorithms to particular datasets. Here the SVM implementation from the LibSVM toolbox (7) was used.

Two different validation approaches were compared for both ankle and wrist data. The first approach was *n-fold cross validation* (*n-fold validation*) (13). In this approach, data (consisting of the windowed sections of data-label pairs) are randomized and divided into *n* different subsets (folds). The algorithm is trained on *n-1* subsets and tested on the remaining one. The second approach was leave-one-subject-out cross validation (LOSO validation) (11). In this case, the subsets correspond to data from the various participants. Recognition

models were trained on data from all subjects except one that is used for the test phase. In both cases the procedure was repeated to test all data. At the end of the procedure, results are aggregated by summing all the obtained confusion matrices. Cross-validation is a well-established technique used in pattern recognition experiments to avoid training and testing on the same data when only small datasets are available (13). The drawback of the *n*-fold validation approach is that temporally-adjacent bits of data may be split into the training and test sets, encouraging the algorithm to overfit the data, inflating positive results. The LOSO validation approach avoids this problem and is more likely to generalize to new data; it is therefore a preferable method.

Results are evaluated in terms of overall accuracy and F1-score for each class. The F1-score

is defined as the harmonic mean of precision and recall, $F1=2\frac{precision\cdot recall}{precision+recall}$ where precision=TP/(TP+FP) and recall=TP/(TP+FN). True positives (TP) are data correctly classified within the selected class. False positives (FP) are those data that are incorrectly classified as belonging to the selected class. False negatives (FN) are data belonging to the selected class that are incorrectly classified in another one. The F1-score merges information about precision and recall in a single number; it ranges from 0 to 1, where 1 is a perfect classification.

Results

As a first step, preliminary activity classification outcomes obtained by testing the algorithm performance with different feature sets (FS) are shown in Table 1. Corrected wrist data were processed using LOSO validation. The effect of discarding features is evaluated: time-frequency analysis (FS #2), wavelet transform (FS #3), mean and standard deviation (FS #4), and maximum and minimum value (FS #5). FS #6 to FS #8 show the effects of discarding sets of the features. The introduction of the "minimum" and "maximum" values improves the classification accuracy by 1.8 percentage points (going from FS #6 to FS #3), whereas removing the wavelet based features reduces the wrist data classification accuracy by 0.6 percentage points. All subsequent results presented here therefore refer to FS #3 that achieves an 84.7% overall accuracy with LOSO validation.

Table 2 shows the results of the activity classification problem in terms of confusion matrices that were obtained after and before the introduction of the ankle data based correction. The amount of ambulation data discarded by this correction is 1.1% of the ambulation data (0.4% of the total). The ankle data based correction was applied to both ankle and wrist datasets to remove windows that were clearly mislabeled as ambulation given the ankle data. Results are shown for both wrist and ankle data, considering the two proposed validation approaches: part A and B contains LOSO validation of uncorrected and corrected data, respectively. Part C shows results obtained with 10-fold validation of corrected data. Results for 10-fold validation of uncorrected data were omitted for brevity. Table 2, part B shows the wrist worn sensor was ~10% less accurate than the ankle-worn sensor, which shows impressive performances on the selected activities with LOSO validation (wrist 84.7%, ankle 95.0%). Most errors for wrist occur in the cycling class and the accuracy of walking detection is 87.2%.

A detailed version of the confusion matrices shown in Table 2 part B is presented in Table 3, and classification accuracies by participant and activity category are reported in figure 1. These data provide insight into the nature of the classification errors, as discussed later.

Time frequency analysis improves with longer windows, but increasing window size reduces the time resolution of outcomes and increases latency in a real-time implementation.

12.8 s windows introduce substantial latency. Table 4 shows LOSO validation results for the ankle and wrist corrected data, varying the size of windows between 12.8 s, 4 s, and 2 s.

For real-time implementation of the algorithm, the computational complexity of feature computation and the classification algorithm must be considered against overall performance. On a 1.6 GHz Intel Centrino2 processor with 4 GB RAM using Mathworks Matlab R2008b on a 64-bit Windows 7 operating system, the algorithm using 12.8 s windows was measured on the dataset to require the following processing time per window: 6.0 ± 0.5 ms (mean \pm standard deviation) to pre-process the window, 104.9 ± 4.0 ms to evaluate features from it, and 0.6 ± 0.1 ms to classify it according to the trained classification rules and parameters. The total time needed for classifying each window is therefore 111.5 ± 4.1 ms. Using shorter windows (4 s) requires 96.8 ± 4.2 ms per window. Introducing the wavelet analysis increases the time by $\sim 12\%$ (13 ms for 12 s windows and 12 ms for 4 s windows, on average).

Discussion

In this work we obtained a classification algorithm which shows good assignment of a wide variety of activities into four distinct classes using raw data from a triaxial accelerometer worn either on the wrist or ankle. The algorithm is computationally-efficient and could be implemented in real-time with short latency to recognize the activity of the user.

In Table 5 our activity classification algorithm is compared to state-of-the-art solutions. The solutions or studies published to date on this topic have four limitations. The first limitation is that the algorithms have been trained and evaluated on small pools of participants with little data per participant (8, 14, 25, 31). A second limitation is that some studies used *n*-fold validation but *not* LOSO validation (8, 25, 32). The third limitation is that some studies used 50% overlapping windows (8, 14, 31). With overlapping windows, some of the same data appears in two windows, potentially inflating recognition results – especially if this overlapping window technique is applied jointly with *n*-fold validation, as in (8).

Another limitation of the prior research, and perhaps the most important, is that the algorithms proposed may have been tested on datasets that fail to represent non-laboratory behavior. The prior studies evaluate algorithms on stereotypical activities performed for only a few seconds. As a consequence, the total amount of classified minutes of wrist data studied, and the total minutes per activity class recognized, is limited, as shown in Table 5. Variability of the behavior within activities lasting for only a few seconds is likely to be reduced, especially for highly-structured behaviors such as treadmill walking. For example, the small postural changes someone might make while sitting in a chair reading or while working on a computer typing are more likely to be recorded if the data logging focuses on activities lasting more than a few seconds. All prior studies used short activities except the sport activities classification by Siirtola et al. (25) in which activities lasted for 10 minutes That study, however, did not include sedentary activities of daily living where wrist sensor data may be difficult to interpret; some are included in our protocol (see Table 3).

Classification rates achieved by Zhang et al. (32) are higher than those reported here. This work uses the LOSO validation instead of the n-fold validation because LOSO is a more realistic test. Further, this work includes cycling data, which appears to be the most confounding activity for wrist-based activity recognition. In Zhang et al., according to the reported confusion matrices (and assuming non overlapping windows), 331 minutes of data were classified. Here, 1609 and 1633 minutes of recordings, for wrist and ankle respectively, were classified. This difference in the number of available windows per class between the

wrist and the ankle sensors is due to gaps in data related to the wireless Bluetooth data transmission.

As expected, activity classification results obtained on our dataset with 10-fold validation are higher (1.9 percentage points for the wrist data) than those using a LOSO validation (see Table 2 part B and C). The 10-fold validation uses a subset of data from all participants for training the model: it is less affected by data inter-subject variability than the LOSO validation approach in which the tested subject is not considered in the definition of classification rules. Using LOSO validation is a better test because it ensures that absolutely no data from an individual has been included in the training set, and therefore bias is not introduced into the testing set.

Tables 2 and 3 highlight some challenges with wrist-based activity recognition:

- Cycling activities are more difficult to detect than the other classes from wrist data, due to the stable position of the wrist on the bike handle (see also the lower F-score values for cycling in Tables 1 and 4). Only the Siirtola et al. (25) prior study included cycling.
- Assuming that natural walking is typically level walking at a self-selected speed (~3 mph), the "treadmill 0% incline 3mph" activity is likely to be similar to natural walking, and ambulation detection is best for these two activities on the wrist. Other variations on walking, such as different speeds and inclinations, increase error rates. This increase may result from having fewer training instances for these walking variations.
- The classifier is capable of recognizing ambulation from wrist data even if the participant is carrying a load, which would alter the nature of the signal. The prior studies do not include examples of carrying objects that might impact ambulatory wrist movement.
- Using ankle data, the most challenging classification category is the "other activities" class. In this case, the participant may perform the sub-activities without moving his or her feet. This is evident for the "painting" sub class, whereas the "sweeping with broom" subclass yields higher classification rates probably because it involves moving the feet.
- Sedentary activities are often classified as "other activities" if the subject is in an
 upright position or if the subject is sorting files or paperwork. This is more
 evident for wrist data, probably because the hand movement in these activities
 can be similar to the "other activities" class.
- Downhill cycling using the ankle sensor location is recognized less well than
 level cycling and exercise bike pedaling. This is probably related to the
 frequency of pedaling, which is lower in uphill conditions and may be zero
 while going downhill. The Siirtola et al. (25) study that included cycling had
 only cycling and spinning activities without any distinction in the incline of the
 path being reported.

Figure 1 shows performance by participant, highlighting person-specific outliers such as the large variability in the "cycling" and "other activities" detection from the wrist data. Figure 1 also shows differences in classification performance between the ankle and wrist data. For example, with cycling wrist classification, it is evident that data from some participants are better classified than others (see for example #2 and #3, and #4 and #5). This high intersubject variability suggests that further improvement may require introduction of a subject-specific adaptation of classification rules, where parameters of the trained classifier are

tailored to the particular participant's behavior with minimal intervention by the user. This subject-specific adaptation, common in fields such as speech recognition, is one of our focuses for future research. In doing so, the critical importance of accurate labeling of activities, discussed above, must be considered.

In this work, we clustered 26 activity types into 4 broad activity classes of ambulation, cycling, sedentary, and other. We also evaluated the same algorithm on the same dataset using the simpler binary classification problem detecting "ambulation" versus "everything else". The LOSO overall accuracy for wrist data rises to 93.5% in this case. This confirms that the recognition of ambulation from wrist data is feasible, even with an algorithm suitable for real-time implementation.

Window length

To better describe typical everyday activities, it would be preferable to limit window length to smaller values. In fact, many posture and ambulation "bouts" are intermixed and last less than 12.8 s. In Table 4 it is confirmed that by reducing the time length of windows from 12.8 s to 4 s, overall classification rates for wrist data slightly decrease. A further reduction to 2 s windows results in a greater decrease in classification rates. The ankle data recognition is only modestly impacted by shorter window lengths. The cycling data for the wrist sensor shows that the highest classification accuracy is obtained with 2 s windows. For some activities such as cycling shorter windows may be preferable because there is less likely to be variability during the window (e.g., stopping pedaling) that may create a misclassified activity due to a transition in the type of movement. Moreover, more windows are available for training if they are shorter, and this may improve classifier performances. The results on this dataset suggest that in real-time systems, window lengths could be reduced from 12.8s to 4s, reducing latency substantially, with little impact on overall performance.

Window length impacts computational efficiency. The presented results were obtained after testing several combinations of feature sets. As shown in Table 1, wavelet transform features slightly improve the classification accuracy. However, the improvement of 0.6 percentage points is at the cost of a 12% increase in computation time, which is substantial when considering real-time implementation on mobile devices. This is the reason why FS #3 from Table 1 was used in this work. The improvement in classification accuracy achieved by introducing the minimum and maximum SMV value features is higher than using the wavelet features. This is not surprising because it is evident from data that the range of recorded SMV of the acceleration differs for different activities and, as Table 1 shows comparing columns FS #2, FS #3, and FS #7, the time-frequency analysis features already provide most of the benefit of the wavelet features.

Generalizability of approach

The proposed methodology could be reproduced on any device that outputs raw 3-axis acceleration data at 40Hz or higher; it is not dependent on the Wockets sensors. Devices with dynamic range other than ± 4 g might require that the algorithms be retrained, or that the raw data be mapped into the ± 4 g range. In this work a 90 Hz sampling rate was used because it was the upper limit of our system and the sampling rate can be downsampled. However, the frequency content of human movement is limited, and tests confirmed that the sampling rate could be reduced to 40 Hz without degrading classification accuracy.

Most of the computational complexity of using SVM classification occurs during the training of the classifier. Once the models are built from the training data, the algorithm can be run efficiently in real-time. The time needed to process a 12.8 s window compared to the case of a 4 s window is slightly higher, even though the complexity of the classification step

in the second case is higher because a larger training dataset may result a larger number of support vectors that must be processed. Therefore, a real-time system using 4 s windows would not only reduce latency, but also overall computational complexity. This would make real-time implementation of activity class detection more computationally feasible on mobile devices.

Transitions crop and ankle-based correction of wrist data

Research assistants were trained to record activities in real time on a portable computing device, but it is still difficult to annotate behavior in real time without making mistakes at some activity transitions. We discarded some windows around activity transitions to minimize data labeling inaccuracies that might impact the results. It is worth noting that the overall LOSO validation accuracy for wrist data does not change significantly using all data without this transition rejection, whereas the ankle data classification accuracy is reduced by 2.3 percentage points when transition data are used. The erroneous classifications of some ambulation data as sedentary activity for wrist data (see Table 2 part A) are resolved by cleaning the data using the ankle-based correction (see Table 2 part B). Using this corrected data also reduces errors classifying sedentary data as ambulation, because after the windows mislabeled as ambulation are removed, overall performance improves. The errors in labeling that we were able to discover by comparing the wrist and ankle data demonstrate the importance of providing the training algorithm a cleanly-labeled dataset – even a small percentage of mislabeled data impacts performance on a small dataset. Despite laboratory conditions and human "gold-standard" direct-observation labeling, we still found errors in our data because of the rapid switching between ambulation and non-ambulation during everyday movement. As new mobile systems are developed that require end-users to provide calibration data, the importance of accurate activity labeling must be considered carefully. If labeling is obtained in less-controlled conditions (e.g., by end-users or the system themselves), errors are likely to be far more common. Using both wrist and ankle sensors in future work may provide one method by which labeled ambulation data may be verified to improve training of wrist-only algorithms.

To make a fair evaluation of classification accuracy possible, all windows containing multiple activities were discarded both from train and test phases. As a consequence, this work establishes an upper bound on the overall performances of the algorithm. In future work we intend to deal with such transition windows in the test phase, providing information on the likelihood of each window being classified in each of the available classes (i.e., a soft-assignment labeling approach). Results could then be compared to the likelihoods of the labels for any given segment; rather than detecting the activity, the algorithm must detect the likelihood of the activity and that is compared to the likelihood of all possible labels. A strong assumption made in this work is that people are not multi-tasking during most activities, and specifically not during the activities we asked them to perform. As a consequence, new algorithms may be necessary to detect complex multi-tasking behaviors. Complex situations like walking and talking on the phone or walking and pushing a stroller are left to future work.

Study strengths and weaknesses

This study has a number of weaknesses related to a relative small and homogenous sample of adults, a selected group of physical activities that account for only a portion of the time and activities many adults perform throughout their day and use of data collected only during "steady-state" activities. Study strengths include data collection according to well-defined and executed protocols, extensive cleaning of data using customized software to insure accuracy, advanced data analytical procedures, an unbiased validation approach and direct comparison of results with recently published data.

Conclusions

As large surveillance studies move towards using wrist-worn accelerometers that collect raw data, the research community will need methods to compute summary statistics on the raw data. Of particular importance is the reliable detection of sedentary vs. ambulatory activity, because there are many sedentary activities that involve relatively large amounts of wrist movement (e.g., animated talking, keyboard typing, repetitive desk work). Ambulatory activities such as walking typically involve repetitive, cyclic motion of various parts of the body, and so intuitively one way to identify such movement from raw accelerometer data at the wrist might be to use frequency-domain features. In addition, we might expect a monitor on the ankle, which would rarely move repetitively and extensively except during ambulation, to perform better than a sensor at the wrist. In this work, we have confirmed this intuition to be true on a dataset of 26 activities collected from 33 people. We find that when trying to detect if a person is engaged in one of four categories of activities – ambulation, cycling, sedentary activities and other activities – a good solution is to use frequency domain features (plus several other simple features) computed on short windows (12.8 s or 4 s) of raw data. On the same dataset, the ankle does outperform the wrist by 10.3%. We find that once frequency domain features are included, the addition of the more computationally complex wavelet features provide only modest improvements that probably do not justify the computational cost, especially when looking toward the future where devices will provide real-time (low-latency) feedback as part of just-in-time interventions.

Acknowledgments

Fahd Albinali and Jason Nawyn provided help with the Wockets sensors and data collection and data cleaning.

Sources of Funding. This study was funded by the National Heart, Lung and Blood Institute, National Institutes of Health award #5UO1HL091737 to the Massachusetts Institute of Technology (Stephen Intille, PI) with a subaward to Stanford University (William Haskell, PI). Mr. Mannini was funded by the Italian Ministry of Education, Universities and Research, MIUR. The present study does not constitute endorsement by ACSM.

References

- Albinali, F.; Intille, SS.; Haskell, W.; Rosenberger, M. Using Wearable Activity Type Detection to Improve Physical Activity Energy Expenditure Estimation; Proc. of the 12th ACM Internat Conf. on Ubiquitous computing; Copenhagen, Denmark. 2010; p. 311-20.
- 2. Atallah L, Lo B, King R, Yang G-Z. Sensor positioning for activity recognition using wearable accelerometer. IEEE Trans Biomed Circuits Sys. 2011; 5(4):320–29.
- 3. Bao L, Intille SS. Activity Recognition from User-Annotated Acceleration Data. Pervasive. 2004; 301:1–17.
- 4. Bhattacharya A, McCutcheon EP, Shvartz E, E GJ. Body acceleration distribution and O2 uptake in humans during running and jumping. J Appl Physiol. 1980; 49(5):881–7. [PubMed: 7429911]
- 5. Bourke K, O'Donovan K, Olaighin G. The identification of vertical velocity profiles using an inertial sensor to investigate pre-impact detection of falls. Medical Eng Phys. 2008; 30(7):937–46.
- 6. Centre UBC. Stockport Cheshire. 2009. Category 2 enhanced phenotyping at baseline assessment visit in last 100-150,000 participants; p. 1-33.
- 7. Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. ACM Trans Intel Syst Tech. 2011; 2(27):1–27.
- 8. Chernbumroong, S.; Atkins, AS.; Yu, H. Activity classification using a single wrist-worn accelerometer; Proc. of the 5th Internat Conf. on Software, Knowledge Information, Industrial Management and Applications; 2011; p. 1-6.
- 9. Degen, T.; Jaeckel, H.; Rufer, M.; Wyss, S. SPEEDY: a fall detector in a wrist watch; Proc. of the 7th IEEE Symp. on Wearable Computing; 2003; p. 184-9.

10. Dongwoo, K.; Kim, HC. Activity energy expenditure assessment system based on activity classification using multi-site triaxial accelerometers; Proc. of the 29th IEEE Internat Engineering in Medicine and Biology Conf.; Lyon, France. 2007; p. 2285-7.

- 11. Esterman M, Tamber-Rosenau BJ, Chiu Y-C, Yantis S. Avoiding non-independence in fMRI data analysis: Leave one subject out. NeuroImage. 2010; 50(2):572–6. [PubMed: 20006712]
- 12. Intille, SS.; Albinali, F.; Mota, S.; Kuris, B.; Botana, P.; Haskell, WL. Design of a wearable physical activity monitoring system using mobile phones and accelerometers; Proc. of the 33rd IEEE Internat Eng in Medicine and Biology Conf.; Boston, MA USA. 2011; p. 3636-39.
- Jain AK, Duin RPW, Mao J. Statistical Pattern Recognition: A Review. IEEE Trans Pattern Anal Mach Intell. 2000; 22(1):4–37.
- Kao, T-P.; Lin, C-W.; Wang, J-S. Development of a portable activity detector for daily activity recognition; Proc of the IEEE Internat Symp on Industrial Electronics; Seoul. 2009; p. 115-20.
- 15. Li Q, Young M, Naing V, Donelan JM. Walking speed estimation using a shank-mounted inertial measurement unit. Journal of biomechanics. 2010; 43(8):1640–3. [PubMed: 20185136]
- 16. Mannini A, Sabatini AM. Gait phase detection and discrimination between walking-jogging activities using hidden Markov models applied to foot motion data from a gyroscope. Gait & Posture. 2012; 36(4):657–61. [PubMed: 22796244]
- 17. Mathie MJ, Coster ACF, Lovell NH, Celler BG. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. Physiol Meas. 2004; 25(2):R1–R20. [PubMed: 15132305]
- 18. Mayagoitia RE, Lotters JC, Veltink PH, Hermens H. Standing balance evaluation using a triaxial accelerometer. Gait Posture. 2002; 16(1):55–9. [PubMed: 12127187]
- 19. Pappas IPI, Popovic MR, Keller T, Dietz V, Morari M. A reliable gait phase detection system. IEEE Trans Neural Syst Rehabil Eng. 2001; 9(2):113–25. [PubMed: 11474964]
- Patel S, Hughes R, Hester T, Stein J, Akay M, Dy JG, Bonato P. A Novel Approach to Monitor Rehabilitation Outcomes in Stroke Survivors Using Wearable Technology. Proc. of the IEEE. 2010; 98(3):450–61.
- 21. Plasqui G, Westerterp KR. Physical Activity Assessment With Accelerometers: An Evaluation Against Doubly Labeled Water. Obesity. 2007; 15(10):2371–9. [PubMed: 17925461]
- 22. Rosenberger M, Haskell WL, Albinali F, Mota S, Nawyn J, Intille SS. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. Med Sci Sports Exer. in press.
- 23. Rueterbories J, Spaich EG, Larsen B, Andersen OK. Methods for gait event detection and analysis in ambulatory systems. Med Eng Phys. 2010; 32(6):545–52. [PubMed: 20435502]
- 24. Sabatini AM. Quaternion-based strap-down integration method for applications of inertial sensing to gait analysis. Med Biol Eng Comput. 2005; 43(2002):94–101. [PubMed: 15742725]
- 25. Siirtola, P.; Laurinen, P.; Haapalainen, E.; Roning, J.; Kinnunen, H. Clustering-based Activity Classification with a Wrist-worn Accelerometer Using Basic Features; Proc. of the IEEE Symp. on Computational Intelligence and Data Mining; Nashville, TN.. 2009; p. 95-100.
- 26. Song, Y.; Shin, S.; Kim, S.; Lee, D.; Lee, KH. Speed Estimation From a Tri-axial Accelerometer Using Neural Networks; Proc. of the 29th IEEE Engineering in Medicine and Biology Society Conf.; Lyon, France. 2007; p. 3224-7.
- 27. Troiano, R.; Mc Clain, J. Objective Measures of Physical Activity, Sleep, and Strength in U.S. National Health and Nutrition Examination Survey (NHANES) 2011-2014; Proc of the 8th Internat Conf on Diet and Activity Methods; Roma, Italy. 2012;
- Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc. 2008; 40(1):181–8. [PubMed: 18091006]
- 29. Vapnik, V. The Nature of Statistical Learning Theory. Springer; New York: 2000. p. 314
- 30. Welk GJ, Blair SN, Wood K, Jones S, Thompson RW. A comparative evaluation of three accelerometry-based physical activity monitors. Med Sci Sports Exerc. 2000; 32(9):S489–97. [PubMed: 10993419]
- 31. Yang J-Y, Chen Y-P, Lee G-Y, Liou S-N, Wang J-S. Activity Recognition Using One Triaxial Accelerometer: A Neuro-fuzzy Classifier with Feature Reduction. Lect Notes Computer Sci. 2007; 4740:395–400.

32. Zhang S, Rowlands AV, Murray P, Hurst TL. Physical activity classification using the GENEA wrist-worn accelerometer. Med Sci Sports Exerc. 2012; 44(4):742–8. [PubMed: 21988935]

33. Zijlstra W, Bisseling RW, Schlumbohm S, Baldus H. A body-fixed-sensor-based analysis of power during sit-to-stand movements. Gait Posture. 2010; 31(2):272–8. [PubMed: 19963386]

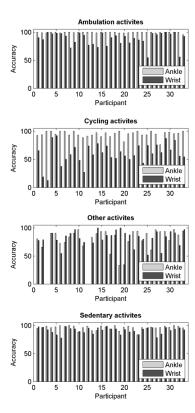


Figure 1. Comparison of wrist/ankle classification accuracies by participant and activity class, showing the large variability in detection of the cycling and other classes from the wrist data.

Table 1

different feature sets. FS#3, which does not include wavelets, provides a good tradeoff between computational cost and overall recognition performance Results classifying 12.8 s windows of activity into four activity classes from wrist data using a SVM recognizer and LOSO cross validation with 10 and is used in the remainder of the paper.

Mannini et al.

Wrist, corrected data	FS #1	FS #2	FS #3	FS #4	FS #5	FS #6	FS #7	FS #8
Time-frequency analysis	`>		`	`>	`	`>		`
Wavelet transform features	`	`		`	`			`
Mean and standard deviation	`	`	`		`	`	`	
Maximum and minimum values	`	`	`	`			`	
Overall accuracy	85.3	82.9	84.7	82.8	83.4	82.9	75.3	81.0
F1-score (walking)	06.0	0.85	0.90	68.0	06.0	0.89	0.79	0.89
F1-score (cycling)	69.0	99.0	99.0	0.61	0.62	0.58	0.43	0.52
F1-score (others)	0.83	0.75	0.82	0.79	0.81	0.81	0.67	0.77
F1-score (resting)	0.88	0.89	0.88	0.85	98.0	98.0	0.84	0.84

Acronyms: FS, Feature Set

Page 15

Table 2

Wrist and ankle classification confusion matrices for the four target activity groups using the SVM classifier with feature set #3. Part A results use LOSO validation with the original data. Part B results use LOSO validation on the dataset where the ankle-based correction ("corrected data") is applied to remove data mislabeled as ambulation. Part C results use the same dataset as in Part B but with 10-fold validation.

Mannini et al.

			Ambulation	Cycling	Other activities	Sedentary
		Ambulation	2206 (85.7 %)	121 (4.7 %)	78 (3.0 %)	168 (6.5 %)
		Cycling	84 (8.1 %)	648 (62.7 %)	16 (1.5 %)	285 (27.6 %)
	Actual label	Other activities	68 (7.1 %)	12 (1.3 %)	773 (81.1 %)	100 (10.5 %)
		Sedentary	33 (1.1 %)	158 (5.3 %)	70 (2.3 %)	2723 (91.3 %)
			Overall accuracy = 84.2 %	· = 84.2 %		
Ankle			Ambulation	Cycling	Other activities	Sedentary
		Ambulation	2550 (96.8 %)	9 (0.3 %)	53 (2.0 %)	23 (0.9 %)
		Cycling	17 (1.6 %)	973 (93.7 %)	20 (1.9 %)	28 (2.7 %)
	Actual label	Other activities	16 (1.6 %)	10 (1.0 %)	794 (80.8 %)	163 (16.6 %)
		Sedentary	2 (0.1 %)	15 (0.5 %)	104 (3.5 %)	2879 (96.0 %)
			Overall accuracy = 94.0%	· = 94.0 %		
art B,	LOSO validat	Part B, LOSO validation, corrected data	В			
Wrist			Ambulation	Cycling	Other activities	Sedentary
		Ambulation	2121 (87.2 %)	112 (4.6 %)	74 (3.0 %)	126 (5.2 %)
		Cycling	86 (8.3 %)	650 (62.9 %)	16 (1.5 %)	281 (27.2 %)
	Actual label	Other activities	59 (6.2 %)	16 (1.7 %)	778 (81.6 %)	100 (10.5 %)
		Sedentary	36 (1.2 %)	156 (5.2 %)	70 (2.3 %)	2722 (91.2 %)
			Overall accuracy = 84.7 %	. = 84.7 %		
Ankle			Ambulation	Cycling	Other activities	Sedentary
		Ambulation	2481 (99.5 %)	10 (0.4 %)	2 (0.1 %)	(%0)0
		Cycling	13 (1.3 %)	975 (93.9 %)	21 (2.0 %)	29 (2.8 %)
	Actual label	Other activities	6 (0.6 %)	12 (1.2 %)	802 (81.6 %)	163 (16.6 %)
		Sedentary	0 (0.0 %)	16 (0.5 %)	105 (3.5 %)	2879 (96.0 %)

Page 16

Page 17

Wrist			Ambulation	Cycling	Other activities	Sedentary
Part C,	10-fold cross-v	Part C, 10-fold cross-validation, corrected data	ed data			
Wrist			Ambulation	Cycling	Other activities	Sedentary
		Ambulation	2196 (90.3 %)	87 (3.6 %)	67 (2.8 %)	83 (3.4 %)
		Cycling	72 (7.0 %)	687 (66.5 %)	17 (1.6 %)	257 (24.9 %)
	Actual label	Other activities	60 (6.3 %)	16 (1.7 %)	791 (83.0 %)	86 (9.0 %)
		Sedentary	31 (1.0 %)	154 (5.2 %)	66 (2.2 %)	2733 (91.6 %)
			Overall accuracy = 86.6 %	' = 86.6 %		
Ankle			Ambulation	Cycling	Other activities	Sedentary
		Ambulation	2486 (99.7 %)	6 (0.2 %)	1 (0.0 %)	0 (0.0 %)
		Cycling	13 (1.3 %)	980 (94.4 %)	22 (2.1 %)	23 (2.2 %)
	Actual label	Other activities	5 (0.5 %)	11 (1.1 %)	823 (83.7 %)	144 (14.6 %)
		Sedentary	0 (0.0 %)	10 (0.3 %)	89 (3.0 %)	2901 (96.7 %)

Acronyms: LOSO, Leave-one-subject-out

Table 3

Wrist and ankle classification details showing category recognition for each specific activity type. All results refer to LOSO validation of corrected data. The numbers indicate the total windows classified into each of the four major categories for each specific activity type (and percentage). Correct classifications are **bold**.

Ambulation Activities				
Ambulation Activities				
walking. carrying-load	322 (98.8 %)	0 (0 %)	3 (0.9 %)	1 (0.3 %)
stairs: inside and down	14 (100 %)	0 (0 %)	0 (0 %)	0 (0 %)
stairs: inside and up	9 (100%)	0 (0 %)	0 (0 %)	0 (0 %)
treadmill: 3.0mph 0% incline	257 (95.5 %)	0 (0 %)	10 (3.7 %)	2 (0.7 %)
treadmill: 3.0mph 6% incline	87 (91.6 %)	4 (4.2 %)	2 (2.1 %)	2 (2.1 %)
treadmill: 3.0mph 9% incline	71 (89.9 %)	4 (5.1 %)	0 (0 %)	4 (5.1 %)
treadmill: 2.0mph 0% incline	282 (73.4 %)	48 (12.5 %)	12 (3.1%)	42 (10.9 %)
treadmill: 4.0mph 0% incline	351 (86.9 %)	28 (6.9 %)	1 (0.2 %)	24 (5.9 %)
walking. natural	803 (94.1 %)	3 (0.4 %)	39 (4.6 %)	8 (0.9 %)
Cycling Activities				
cycling: 70rpm. 50W. 0.7kg	30 (4.7 %)	393 (61.8 %)	10 (1.6 %)	203 (31.9 %)
cycling: outdoor. level	10 (7.1 %)	105 (74.5 %)	0 (0 %)	26 (18.4 %)
cycling: outdoor. uphill	26 (18.1 %)	102 (70.8 %)	5 (3.5 %)	11 (7.6 %)
cycling: outdoor. downhill	6 (5.4 %)	87 (77.7 %)	2 (1.8 %)	17 (15.2 %)
Other activities				
painting: roller	11 (4.6 %)	1 (0.4 %)	206 (86.2 %)	21 (8.8 %)
painting: brush	7 (3.2 %)	1 (0.5 %)	192 (86.9 %)	21 (9.5 %)
sweeping with broom	42 (8.5 %)	14 (2.8 %)	393 (79.7 %)	44 (8.9 %)
Sedentary activities				
sitting at computer: internet search	2 (0.5 %)	29 (7.3 %)	2 (0.5 %)	366 (91.7 %)
sitting at computer: typing	0 (0 %)	15 (12.1 %)	0 (0 %)	109 (87.9 %)
sitting: writing	4 (0.9 %)	31 (7.0 %)	0 (0 %)	411 (92.2 %)
sitting: reading	3 (0.7 %)	24 (5.3 %)	4 (0.9 %)	426 (93.2 %)
sorting files or paperwork	3 (0.8 %)	27 (7.5 %)	30 (8.4 %)	298 (83.2 %)
lying: on back	3 (0.7 %)	4 (1.0 %)	0 (0 %)	409 (98.3 %)
lying on left side	0 (0 %)	0 (0 %)	0 (0 %)	26 (100 %)
lying on-right-side	0 (0 %)	1 (2.0 %)	0 (0 %)	48 (98.0 %)
sitting: legs straight	7(1.7 %)	10 (2.4 %)	10(2.4 %)	393 (93.6 %)
standing still	9 (3.1 %)	13 (4.5 %)	20 (6.9 %)	247 (85.5 %)
Ankle,	Ambulation	Cycling	Other Activities	Sedentary
Ambulation Activities				
walking. carrying-load	340 (99.7 %)	1 (0.3 %)	0 (0 %)	0 (0 %)
stairs: inside and down	22 (100 %)	0 (0 %)	0 (0 %)	0 (0 %)
stairs: inside and up	9 (100 %)	0 (0 %)	0 (0 %)	0 (0 %)

Mannini et al.

standing still

Wrist, Ambulation Cycling Other Activities Sedentary treadmill: 3.0mph 0% incline 285 (99.7 %) 1 (0.3 %) 0(0%)0(0%)treadmill: 3.0mph 6% incline 0 (0 %) 0 (0 %) 0 (0 %) 113(100 %) treadmill: 3.0mph 9% incline 79 (100 %) 0 (0 %) 0 (0 %) 0 (0 %) treadmill: 2.0mph 0% incline 390 (99.5 %) 1 (0.3 %) 1 (0.3 %) 0(0%)treadmill: 4.0mph 0% incline 404 (100 %) 0 (0 %) 0 (0 %) 0 (0 %) 844 (99.6 %) 0 (0 %) walking. natural 3 (0.4 %) 0 (0 %) **Cycling Activities** cycling: 70rpm. 50W. 0.7kg 3 (0.5 %) 632 (99.5 %) 0(0%)0 (0 %) cycling: outdoor. level 5 (3.4 %) 131(90.3 %) 4 (2.8 %) 5 (3.4 %) cycling: outdoor. uphill 3 (2.2 %) 129 (92.8 %) 4 (2.9 %) 3 (2.2 %) cycling: outdoor. downhill 2 (1.7 %) 15 (12.6 %) 88 (73.9 %) 13 (11.8 %) Other activities painting: roller 0(0%)2(0.8%)191 (78.3 %) 51 (20.9 %) 73 (32.0 %) painting: brush 0(0%)2 (0.9 %) 153 (67.1 %) 5 (1.0 %) 20 (3.9 %) sweeping with broom 7 (1.4 %) 479 (93.7 %) Sedentary activities sitting at computer: internet search 0(0%)1(0.3 %) 11 (2.8 %) 379 (96.9 %) sitting at computer: typing 0 (0 %) 1(0.8 %) 3 (2.3 %) 124 (96.9 %) sitting: writing 0(0%)0 (0 %) 9 (2.0 %) 438 (98.0 %) sitting: reading 0(0%)1 (0.2 %) 17 (3.7 %) 444 (96.1 %) sorting files or paperwork 0(0%)2 (0.6 %) 16 (4.5 %) 339 (95.0 %) 0 (0 %) 3 (0.7 %) lying: on back 0(0%)420 (99.3 %) lying on left side 0(0%)0(0 %)1 (2.5 %) 39 (97.5 %) lying on-right-side 0(0%)0(0%)0 (0 %) 56 (100 %) sitting: legs straight 0(0%)2 (0.5 %) 9 (2.3 %) 381 (97.2 %)

0 (0 %)

3 (1.0 %)

20 (6.6 %)

281 (92.4 %)

Page 19

Table 4

Mannini et al.

Results using the SVM classifier on the corrected data and FS#5 with three different sizes of windows.

	Wrist			Ankle		
	12.8 s 4 s	4 s	2 s	12.8 s	4 s	2 s
Overall accuracy (%)	84.7	84.2	82.8	95.0	94.6	93.8
Accuracy (ambulation)	87.2	87.4	87.1	5.66	99.1	98.8
Accuracy (cycling)	67.9	65.2	65.4	93.9	94.2	93.7
Accuracy (others)	81.6	77.0	72.6	81.6	78.7	76.7
Accuracy (sedentary)	91.2	9.06	88.9	0.96	96.1	95.3
F1-score (ambulation)	0.90	0.90	0.89	0.99	0.99	0.99
F1-score (cycling)	99.0	0.68	99.0	0.95	0.95	0.95
F1-score (others)	0.82	0.78	0.74	0.84	0.81	0.79
F1-score (sedentary)	0.88	0.87	0.86	0.95	0.95	0.94

Page 20

Table 5

Recent results using raw data for activity classification from the wrist, as compared to the approach presented in this study. Details, in particular related to potential weaknesses of the validation and windowing approach used in some studies, are discussed in the text.

	Yang et al.	Kao et al.	Siirtola et al.	Chernbumroong et al.	Zhang et al.	This study
# participants	7	4	10	7	09	33
Classes	Sitting Standing PC working Brushing teeth Scrubbing Vacuuming Walking Running	PC working Brushing teeth Knocking a table Hitting Swinging a hand Walking Running	Nordic walking Team sports Spinning, cycling Gym training Roller skating Racket sports Aerobics Walking Running	Sitting Standing Walking Running	Sedentary Household Walking Running	Sedentary Other activities Walking Cycling
Overall accuracy	92.6 %	94.7 %	85.0 %	94.1%	97.0 %	84.2%
Walking accuracy	% 5.96	96.4 %	% 001	Not specified	97.5 %	87.2 %
Sedentary accuracy	94.8 %	93.3 % ‡	Not available	Not specified	99.4 %	% 6:06
Validation approach	OSOT	1 for training, 3 for testing	10-fold	5-fold	10-fold	ОЗОТ
Windowing approach	5.24s windows, 50% overlapping	0.32 s windows, 50% overlapping	100Hz data, down- sampled by compression algorithm	3.88s windows, 50% overlapping	12.8s windows,	12.8s non overlapping windows
Total amount of classified minutes	110	9	Not specified $^{\!$	35*	331	1609 (wrist) 1633 (ankle)
Duration of activities / consecutive time considered (min)	2	0.5	10	1 **	2	Mean 4.7 SD 3.3 Range 0.2 to 16.0
Features	Mean, Variance, Correlation, MAD, RMS, Energy, Inter quartile range	Mean, SD, Variance, Correlation, MAD, RMS	Mean, Variance, Average change in y values between two sequential points, Approximation of S – percentile with S = 1, 2,, 15, 85, 86,, 99	Mean, SD, Variance, Correlation Difference x, y, Coefficient sum, Spectral energy, Spectral entropy Minimum, Difference x, y, z	Mean, SD, Time-frequency analysis features, see text Two time domain features from Wavelet transform,	Mean, SD, Time-frequency analysis features, see text, Maximum value, Minimum value,
Classifier	Neuro-fuzzy (92.6%)	Fuzzy basis function(94.7%)	Clustering + decision tree (85.0%)	Decision tree (94.1%) Neural nets (90.4 %)	Decision tree(97.0%) SVM, linear (96.8%)	SVM, RBF(84.2%)

	Yang et al.	Kao et al.	Siirtola et al.	Chernbumroong et al. Zhang et al.	Zhang et al.	This study
# participants	7	4	10	7	09	33
					Naïve Bayes (95.3%) Logistic regr. (95.7%) Neural Nets (96.8%)	
Computational complexity notes				Requires time- frequency analysis, O(N logN) ***	Requires time- frequency analysis, O (N logN) and discrete wavelet transform, O (N)	Requires time- frequency analysis, O (N logN)

Acronyms and abbreviations: LOSO, leave-one-subject-out; SVM, support vector machines; RBF, radial basis function kemel; RMS, root mean square; SD, standard deviation; MAD, mean absolute deviation.

 $[\]slash\hspace{-0.4em}{}^{\slash\hspace{-0.4em} T}_{\slash\hspace{-0.4em}}$ Result obtained excluding walking and running classes

^{*}Confusion matrix reported in percentage values: minute value not evincible. Each of the subjects performed one test in at least one of the activities lasting 10 minutes.

^{*} Confusion matrix not available. 35 minutes are the total amount of acquired data. We assume that they are all classified.

^{**}Value estimated from the amount of available minutes of data: the right value is not specified in the text and not evincible from available results.

^{****}O(·) is order of growth of an algorithm. That is, how the algorithm's worst-case performance changes as the size N of the data set it operates on increases.