8-31-2011

MOVELETS: A DICTIONARY OF MOVEMENT

Jiawei Bai
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics

Jeff Goldsmith
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics

Brian Caffo
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics

Thomas A. Glass
Johns Hopkins Bloomberg School of Public Health, Department of Epidemiology

Ciprian M. Crainiceanu
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, ccrainic@jhsph.edu

Movelets: A Dictionary of Movement

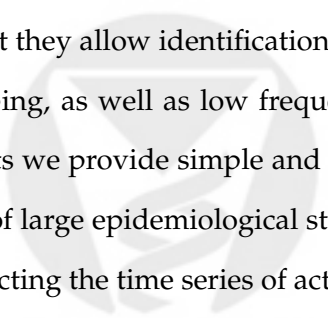Jiawei Bai[a], Jeff Goldsmith[a], Brian Caffo[a],

Thomas A. Glass[b] and Ciprian M. Crainiceanu[a]

August 30, 2011

Abstract

Recent technological advances provide researchers a way of gathering real-time information on an individual's movement through the use of wearable devices that record acceleration. In this paper, we propose a method for identifying activity types, like walking, standing, and resting, from acceleration data. Our approach decomposes movements into short components called "movelets", and builds a reference for each activity type. Unknown activities are predicted by matching new movelets to the reference. We apply our method to data collected from a single, three-axis accelerometer and focus on activities of interest in studying physical function in elderly populations. An important technical advantage of our methods is that they allow identification of short activities, such as taking two or three steps and then stopping, as well as low frequency rare activities, such as sitting on a chair. Based on our results we provide simple and actionable recommendations for the design and implementation of large epidemiological studies that could collect accelerometry data for the purpose of predicting the time series of activities and connecting it to health outcomes.

[a]Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA
[b]Department of Epidemiology, Johns Hopkins University, Baltimore, MD 21205, USA

1

1 Introduction

Accurate measurement of physical activity is necessary for understanding the complex relationship between an individual's health outcomes and his or her behavior profile. Unfortunately, standard measures of activity such as questionnaires and diaries are based on self-report and are subject to known shortcomings. Moreover, these measures typically offer snapshots of activity and do not reflect the dynamic nature of movement in the real world. Recently, progress in sensor technologies and wearable computing devices have allowed researchers to collect real-time information on movement through the use of accelerometers. In this paper, we propose a method for predicting activity types, such as walking, standing and sitting, from a multichannel accelerometer designed with widespread deployment in observational studies in mind.

The use of accelerometers to collect activity information in large-scale observational studies began with the addition of the ActiGraph accelerometer to the National Health and Nutrition Examination Survey (NHANES) in 2003 (Troiano et al., 2008). Early public health work has focused on the quantification of total energy expenditure (Welk et al., 2000; Bussmann et al., 2001; Atienza and King, 2005; Ravi et al., 2005; Boyle et al., 2006; Pärkka et al., 2006; Grant et al., 2006; Ermes et al., 2008; Grant et al., 2008). However, these devices offer the potential to assess more complex questions regarding real-world function and more refined measures of specific activity types. Currently, function is assessed using measures of activities of daily living that depend on retrospective self-report, despite well-documented and substantial measurement error associated with these instruments (Feinstein et al., 1986; McDowell and Newell, 1987). The development of a method to accurately identify activity types and durations based on accelerometer data could alleviate many of the problem associated with self-reported activity data. This is particularly important in the study of aging populations, both because issues with recall are more severe and because understanding physical activity is central to the study of elderly populations in public health (Pate et al., 1995).

We base activity prediction on the idea that movements can be understood in terms of smaller components, which we dub "movelets". Briefly, given accelerometer time series data, we decompose movements into short overlapping segments; these movelets are the elements which make

up motions and activities. Using data with known activity labels, movelets are organized by activity type into "chapters", or collections of movelets with the same activity label. Predictions of unknown activity labels are made by finding the closest match, defined in terms of squared error for all acceleration channels, of an unlabeled movelet to those in chapters. Thus we build our method on the intuition that movements with elements that look similar are likely to have the same labels.

Our data are generated using a single accelerometer positioned on the subject's hip. The accelerometer is built on core chip MMA7260Q by *Freescale*™, and records acceleration in three mutually orthogonal directions for a wide range of sampling frequencies (time points per second) and sensitivities (acceleration per unit of scale). Data were collected during in-laboratory sessions in which subjects performed a collection of activities, including resting, walking, and lying. We observe data for two subjects with two laboratory visits each. Sessions lasted roughly 15 to 20 minutes, and in that time each activity was replicated up to three times. Both the data collection device and activities performed are compatible with the needs of observational studies, especially of elderly populations: the single accelerometer worn at the hip is unobtrusive and wearable in real time, and the activities provide a useful understanding of physical movement. During the data collection, an observer recorded activity start and stop times to provide a time series of movement labels that accompanies the accelerometer signal.

The accelerometer output consists of 3 voltage time series, which are proxy measures of acceleration. The time series vary by amplitude, frequency and correlation along the time course of the corresponding activities. For example, Figure 1 displays two segments of accelerometer data. In the first segment, the subject stands, walks twenty meters, and stands. In the second segment, the subject performs two replicates of lying down and standing up; during each replicate, the subject lies from a standing position, rests for several seconds in the lying position, and rises to a standing position. Three acceleration channels or axes are shown, and activity labels are provided. From this Figure, we see that active periods, in which the subject is walking, rising or lying down, have higher variability than inactive periods, in which the subject is resting in either the standing or lying position. Walking is characterized by periodic acceleration patterns for each axis, although there are differences in amplitude between axes. Replicates of the "Chair
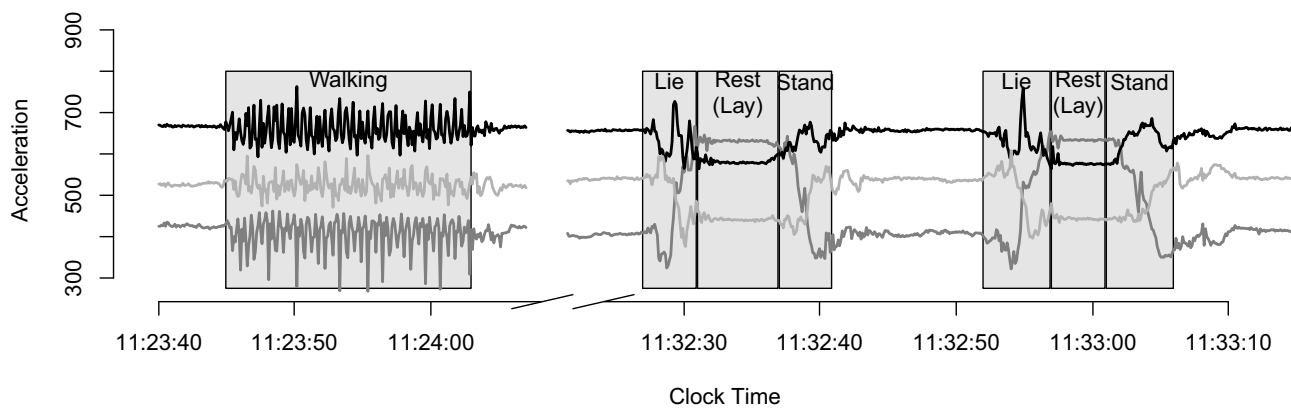
Figure 1: Two segments of accelerometer data. First, a subject walks for approximately 20 second; then, a subject preforms two replicates of "Lie down / Rest / Stand up". Acceleration in three mutually orthogonal directions is shown, and activity labels are included.

Stand" activity display similar patterns, bolstering the intuition that movements that share a label also appear similar visually. Although there are two types of inactivity (standing and lying), the acceleration time series corresponding to these two periods are characterized by low variation around stable constants; however, the ordering and relative position of the axes are different, due to a change in the orientation of the accelerometer with respect to Earth's gravity.

Prediction of physical activity intensity and type has been under intense methodological development in electronic engineering and computer science, but to a lesser extent in statistics. Several prediction methods using either raw or transformed accelerometer data exist, including "cut-point" or linear regression (Freedson et al., 1998; Hendelman et al., 2000), quadratic discriminant analysis (Pober et al., 2006), artificial neural networks (Kiani et al., 1997, 1998; Zhang et al., 2003, 2004; Staudenmayer et al., 2009), Markov Models (Krause et al., 2003; Pober et al., 2006), unsupervised learning (Nguyen et al., 2007) and combined methods (Ravi et al., 2005; Bai, 2011). Previous work has often focused on activity types that are not of interest in public health studies, such as typing and brushing of teeth, or has included multiple accelerometers placed at several locations on a subject's body (Kiani et al., 1997, 1998; Mantyjarvi et al., 2001). A comparison of recent approaches is applied to data generated using five biaxial accelerometers in Bao and Intille (2004). However, these approaches are unsuitable for application to accelerometer data in public health studies, either because they require more sensors than are feasible or

because they are not designed to detect short-term activities like standing from a lying position.

Our approach and taxonomy are inspired by the speech recognition literature, where words or parts of words are matched to known speech patterns. However, the parallel with speech recognition should not be overstated given the large differences between the two activities and measurement instruments. First, speech is often recorded at much higher frequencies (between 8 and 16kHz) than acceleration (10Hz in our dataset), providing density and detail to voice recognition data (Picone, 2005). Second, audio data is inherently single-channel while acceleration is understood in three orthogonal directions, increasing the dimension of the activity prediction problem. In natural speech most sounds and many full words are repeated often, providing an ample training set on which to build a prediction algorithm. In activity prediction, movements can be rarely performed and infrequently observed, making the definition of a training set challenging. Moreover, high fidelity audio recorders could be though lossless reproductions of the original signal. In contrast, accelerometers are weak proxies for activities that are complex and could be ambiguous.

The remainder of the paper is organized as follows. In Section 2 we describe the movelet-based approach to predicting activity based on accelerometer data. Section 3 details the application of our proposed method to the real data described above. We close with a discussion in Section 4.

2 Methods

To predict activities based on accelerometer data, we first define a movelet as a basic element of 3-axis time series data. Collections of movelets paired with known labels (annotations) form chapters, which are in turn organized into reference dictionaries of known movelets and their associated activities. Classification of accelerometer data with unknown activity annotations is based on decomposing the unlabeled data into component movelets, and then matching each unlabeled movelet to these chapters. The label of the best matched chapter is used as a preliminary prediction of the activity of the unlabeled movelet.

2.1 Definitions

We observe data that is a collection of three time series representing the acceleration in three mutually orthogonal axes. Though we have two subjects and each with two visits, we actually treat them as 4 independent visits. Thus denote the data by $\mathbf{X}_i(t) = \{X_{i1}(t), X_{i2}(t), X_{i3}(t)\}$, $t = 1, 2, ..., T_i$, where T_i is the length of the accelerometer time series for visit i. Define an activity label time series $L_i(t)$ such that $L_i(t)$ is a function mapping t to $\{\text{Act}_1, \text{Act}_2, \ldots, \text{Act}_A\}$, $t = 1, 2, \ldots, T_i$, where Act_a denotes activity type a. Let $\mathbb{T}_i$ and $\mathbb{V}_i$ be a partition of observation time for visit i into training and validation sets, respectively. Thus if $t \in \mathbb{T}_i$, then $\mathbf{X}_i(t)$ belongs to the training dataset and has a known activity label $L_i(t)$; otherwise $L_i(t)$ is unknown and is to be estimated. Training sets contain continuous segments or blocks of time to include full examples of each movement type.

Next we define movelets as elements of time series that characterize movement in temporal windows with length H. More specifically, let

$$M_i(t) = \{\mathbf{X}_i(t), \mathbf{X}_i(t+1), \ldots, \mathbf{X}_i(t + H - 1)\},$$

define the movelet at time $t \in \{1, 2, \ldots, (T_i - H + 1)\}$. Note that movelets are made up of time series for all axes of the accelerometer output, and summarizes the pattern of acceleration recorded from time t to $t + H - 1$. The dimension of the movelet $M_i(k)$ is $3H$, because there are 3 concatenated time series, and contains all the accelerometry information for a window of movement of length $H/10$, because time is expressed in 10 Hz. Movelets $M_i(t)$ with $t \in \mathbb{T}$ are paired with their known activity labels and collected into activity-specific "chapters". Thus, we define a chapter $\mathcal{C}_a$ as a collection of movelets $\{M_i(t) : L_i(t) = \text{Act}_a\}$ that share a common label. An important characteristic of movelets is that they overlap; in fact $M_i(t)$ and $M_i(t + 1)$ overlap everywhere, except at time t and $t + H$. This is an important characteristic when there is uncertainty on where the activity actually starts. This happens to be a serious problem even with the best in-lab human annotation. One chapter is constructed for each activity type; chapters are then combined to form a subject-visit specific "dictionary" of movelets and their labels. Dictionaries are distinct for subjects and visits to control for differences between the

movement patterns for different subjects and to account for changes in the orientation of the accelerometer at different visits. This dictionary is used as a reference for movelets $M_i(t)$ with $t \in \mathbb{V}$. Table 1 displays an example of a subject-specific dictionary consisting of A chapters in total. Each chapter is constructed using the training set and is made up of movelets, the short components of three-axis accelerometer data.
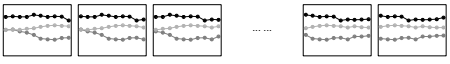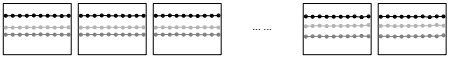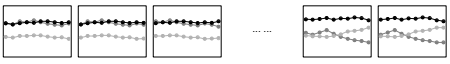
Dictionary

Chapter	Activity	Movelets	
$\mathcal{C}_1$	Activity 1	$\{M_i(t) : L_i(t) = \text{Act}_1\}$	
$\mathcal{C}_2$	Activity 2	$\{M_i(t) : L_i(t) = \text{Act}_2\}$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\mathcal{C}_A$	Activity A	$\{M_i(t) : L_i(t) = \text{Act}_A\}$	

Table 1: A subject-specific dictionary with with A chapters, one for each activity type. Each chapter consists of movelets, short overlapping segments of three-axis accelerometer data, which are illustrated in the far-right column of the table.

The definitions of movelets, chapters, and dictionaries given above provide a useful analogy for our proposed classification method. Given unlabeled accelerometer data that has been decomposed into movelets, we use the dictionary as a reference by "looking up" an unlabeled movelet and finding its best match among known movelets. The label associated with the best match, which is the chapter title, is used to predict the unknown label. Matching, which is described below, quantifies the intuition that movelets with similar visual appearances are likely to be components of the same larger movement.

2.2 Matching and Labeling

Given an unlabeled movelet $M_i(t_0)$, we predict the label $L_i(t_0)$ first by matching $M_i(t_0)$ to a chapter in the dictionary described above. To be more specific, the closest match for movelet $M_i(t_0)$ in the dictionary is $M_i(t')$, where

$$t' = \operatorname*{argmin}_{t \in \mathbb{T}} \left[D\{M_i(t), M_i(t_0)\} \right].$$
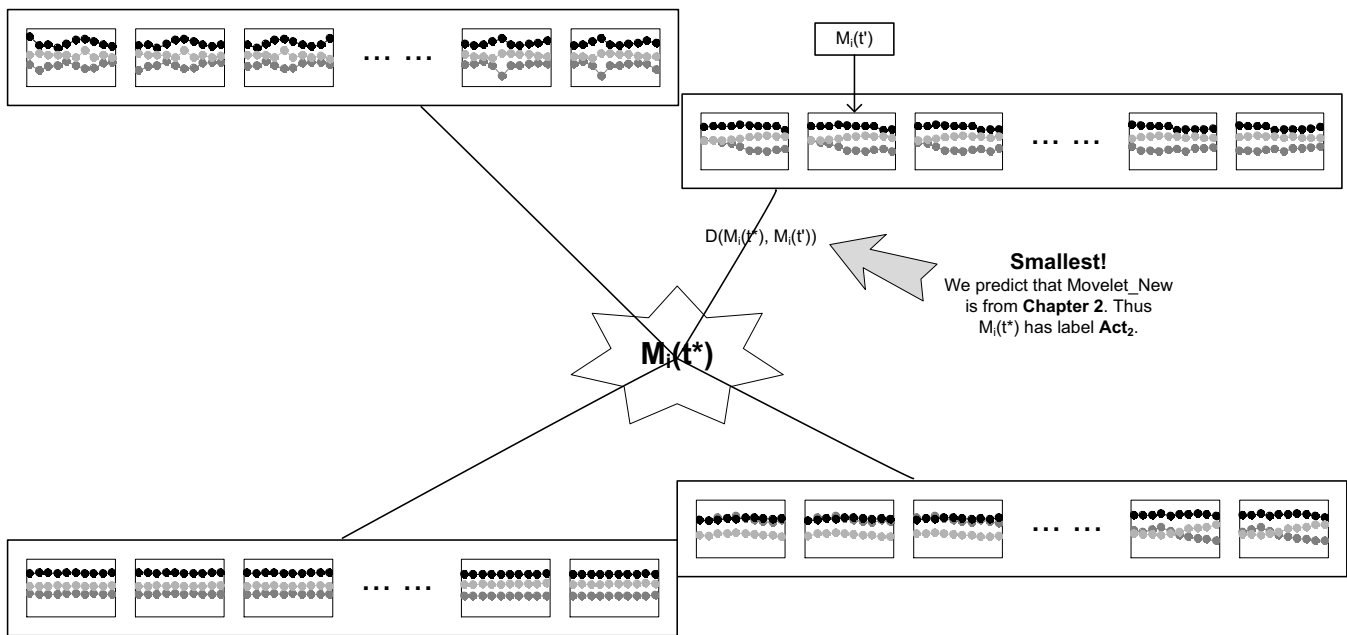
Figure 2: A display of matching an unlabeled movelet $M_i(t^*)$ to 4 chapters in the dictionary. Points in each chapter represent labeled movelets corresponding to the activity associated with this chapter. The distance between the unlabeled $M_i(t^*)$ and each chapter is given by the minimum distance between $M_i(t^*)$ and the movelets in each chapter. After $M_i(t^*)$ is compared to all reference movelets in the dictionary, it is matched to Chapter 2 which provides the smallest distance among all the 4 chapters.

The distance function $D(\cdot, \cdot)$ is

$$D[M_i(t_1), M_i(t_2)] = \frac{1}{3} \sum_{p=1}^{3} \sqrt{\sum_{h=1}^{H} [X_{ip}(t_1 - 1 + h) - X_{ip}(t_2 - 1 + h)]^2}. \tag{1}$$

Thus, distance between movelets averages the difference taken over all acceleration axes. Based on this match, an estimate for the unknown label is $L_i^*(t_0) = L_i(t')$; that is, we take the label associated with the best dictionary match and use it to estimate the unknown label. Figure 2 gives a schematic of the matching process, in which an unlabeled movelet $M_i(t^*)$ is compared to a dictionary with 4 chapters. The distance between $M_i(t^*)$ and all reference movelets is calculated using the distance function (1). After $M_i(t^*)$ is compared to all reference movelets in the dictionary it is matched to Chapter 2, because movelet $M_i(t')$ in Chapter 2 along with $M_i(t^*)$ provides the smallest distance.

After preliminary labels $L_i^*(t), t \in \mathbb{V}$, are generated using the matching step, a majority voting

procedure is used to select final estimated labels $\hat{L}_i(t)$. Each element of $\{L_i^*(t), L_i^*(t+1), \ldots, L_i^*(t+H)\}$ ($t \in [0, T_i - H]$) is considered a single vote, and the activity with the most votes in this set is the estimate $\hat{L}_i(t)$. An advantage of this procedure is that it smooths the predicted labels $\hat{L}_i(t)$ by taking into account the fact that movements are continuous, meaning that neighboring movelets contain information about the current activity.

2.3 Movement Fingerprints and Lazy Movelets

To increase the accuracy of our dictionary-based classification method and decrease the computational burden of the looking-up process, each chapter must be carefully constructed to include useful information while excluding redundant or less useful movelets. With this in mind, chapters that were built in the manner described above can be fine-tuned using the identification of what we will label "fingerprint" and "lazy" movelets.

First, each chapter must to include the signature movelets of the corresponding activity. We refer to these defining movelets as "fingerprints" because they provide excellent prediction of a specific activity related to the chapter. Fingerprints are thus the characteristic acceleration time series associated with a movement, and are most often used when matching new movelets of the same activity. Second, unnecessary or redundant information should be removed from the chapter. For example, a chapter built on several seconds of walking will include many near-identical movelets due to the periodic nature of the activity. Further, there often exist "lazy" movelets which, contrary to fingerprints, are not commonly matched to and do not usefully identify the activity; rather than aiding prediction, these can be falsely matched to by movelets of other activities. Both redundant and lazy movelets can be excluded from a chapter to increase computational performance and reduce the number of errors. Finally, some movements share very similar movelets. These "ambiguous" movelets can lead to misclassification due to very close matches in multiple chapters. In this situation, an ambiguous movelet can be removed from one chapter so that matches will be made to the remaining movelet; the choice of which movelet to retain will depend on the relative importance of correctly classifying the two movements.

As an example of both fingerprints and lazy movelets, Figure 3 displays the chapter for

An example of Chapter RS_Stand

Lazy Movelets

Chapter
RS_Stand

No.1
No.5
No.9
No.13

11:32:26   11:32:30   11:32:34   11:32:38   11:32:42

| Movelet No.1 | Movelet No.2 | Movelet No.3 | Movelet No.4 | Movelet No.5 | Movelet No.6 |

| Movelet No.7 | Movelet No.8 | Movelet No.9 | Movelet No.10 | Movelet No.11 | Movelet No.12 |

| Movelet No.13 | Movelet No.14 | Movelet No.15 | Movelet No.16 |
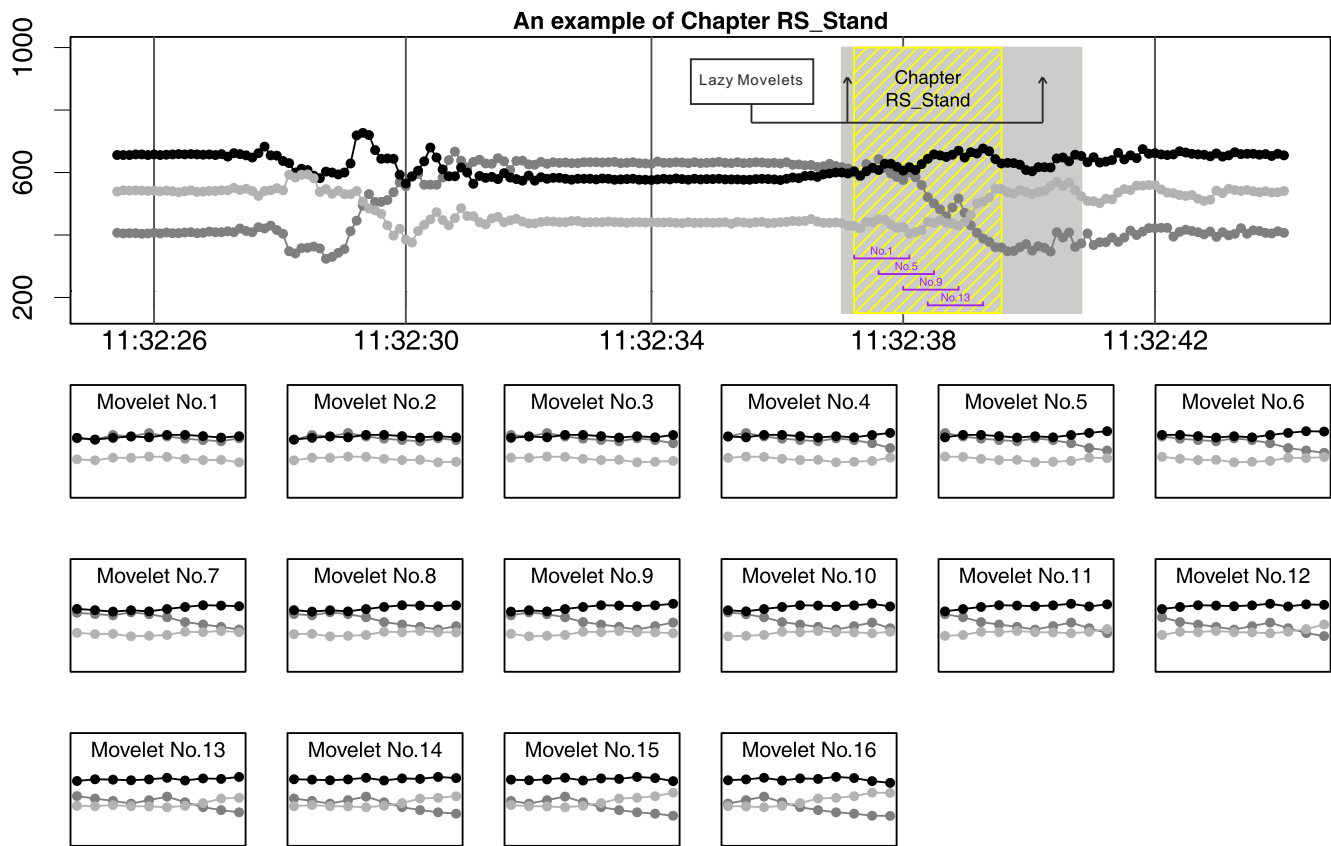
Figure 3: The chapter "Standing from Lying", which consists of 16 movelets. In dark grey is the section of the acceleration data used to construct the chapter; in light grey are time points with the same activity label, but that are excluded from the chapter as "lazy" movelets.

"Standing from Lying" from a movelet dictionary. We used only the region shown in dark gray to construct the chapter, despite the fact that the areas shown in light gray are also labeled by a human observer as "Standing from Lying". The fingerprint of this activity is the pattern that the red time series goes down while the green one goes up. The movelets in the light gray bands are lazy movelets, and do not distinguish this activity from others. We removed the lazy movelets from the annotated time period and built the library conservatively to make the chapter a more useful reference for future unlabeled activities.

2.4 Summary

Movelet-based analysis of accelerometer data is built on the intuition that movements with similar acceleration patterns at the elemental level are likely to be generated by the same activity. Using this idea, we decompose movements into overlapping segments and construct reference chapters and dictionaries; given unlabeled time series, we match to the reference and use the best match to predict the unknown activity type. Movement fingerprints are identified to strengthen the construction of chapters and to aid in the basic understanding of movements, while lazy movelets are eliminated to reduce classification error and computation time. The result is a conceptually clear method for activity prediction that is computationally feasible and scalable to large datasets.

3 Application to LIFEmeter Data

We now apply our methods to data from two subjects, each with two visits. Data were collected in the development of the LIFEmeter multi-sensor device, intended to assess physical function in large-scale observational studies. Subjects were observed in a clinical setting, and performed physical activities that are common in daily living. The following activities were selected as important in understanding physical function in real-world setting: walking, standing from sitting, standing from lying, sitting from standing, and lying from standing. Three sedentary states (standing, sitting, and lying) were also collected. Table 2 lists all activities observed and provides abbreviations that will be used through the remainder of this Section.

An observer annotated the time points at which an activity was started and completed, providing activity labels $L_i^{obs}(t)$. Annotations were imperfect due to early or late start and stop points, to rounding times to the nearest second, and to misalignment. Obvious errors in the observed labels were detected and corrected through comparison with the accelerometer output to create labels used to construct movelet dictionaries and assess the predictive performance of our algorithm.

Activity List	
Activity	**Alias**
Rest (Stand)	Rest1
Rest (Sit)	Rest2
Rest (Lie)	Rest3
Standing from Chair	CS_Stand
Sitting Down from Standing	CS_Sit
Lying Down from Standing	RS_Lie
Standing from Lying	RS_Stand
Walking	Walk

Table 2: A list of activities of interest, with abbreviates used in remaining Figures and text.

3.1 Constructing the dictionary

Following the method described in Section 2, we build a dictionary with 8 chapters of activities for each subject and visit. First, we partitioned the accelerometer data into training and validation sets $\mathbb{T}_i$ and $\mathbb{V}_i$. Using the training set, we decompose movements into movelets and organize by activity type. For activities with well-defined beginnings and endings, such as "CS_Stand" and "CS_Sit" , we use the first replicate as training data and reserve the remaining replicates as testing data. Chapters for these activities contain between 5 and 30 movelets each, depending on the duration of the activity. For continuous movements that lack well-defined beginnings and endings, such as walking or resting, we extract segments lasting 2 to 3 seconds that are clearly labeled with a particular activity to build the corresponding chapter. This is done to prevent chapters from becoming too large, and, since these activities are periodic, to prevent redundant information from being included in the reference.

3.2 Initial Results

After constructing dictionaries for each subject and each visit using the training data, we predict activity labels $\hat{L}_i(t)$ for $s \in \mathbb{V}_i$ by matching movelets to the reference and implementing the majority voting step. Figure 4 details this analysis.
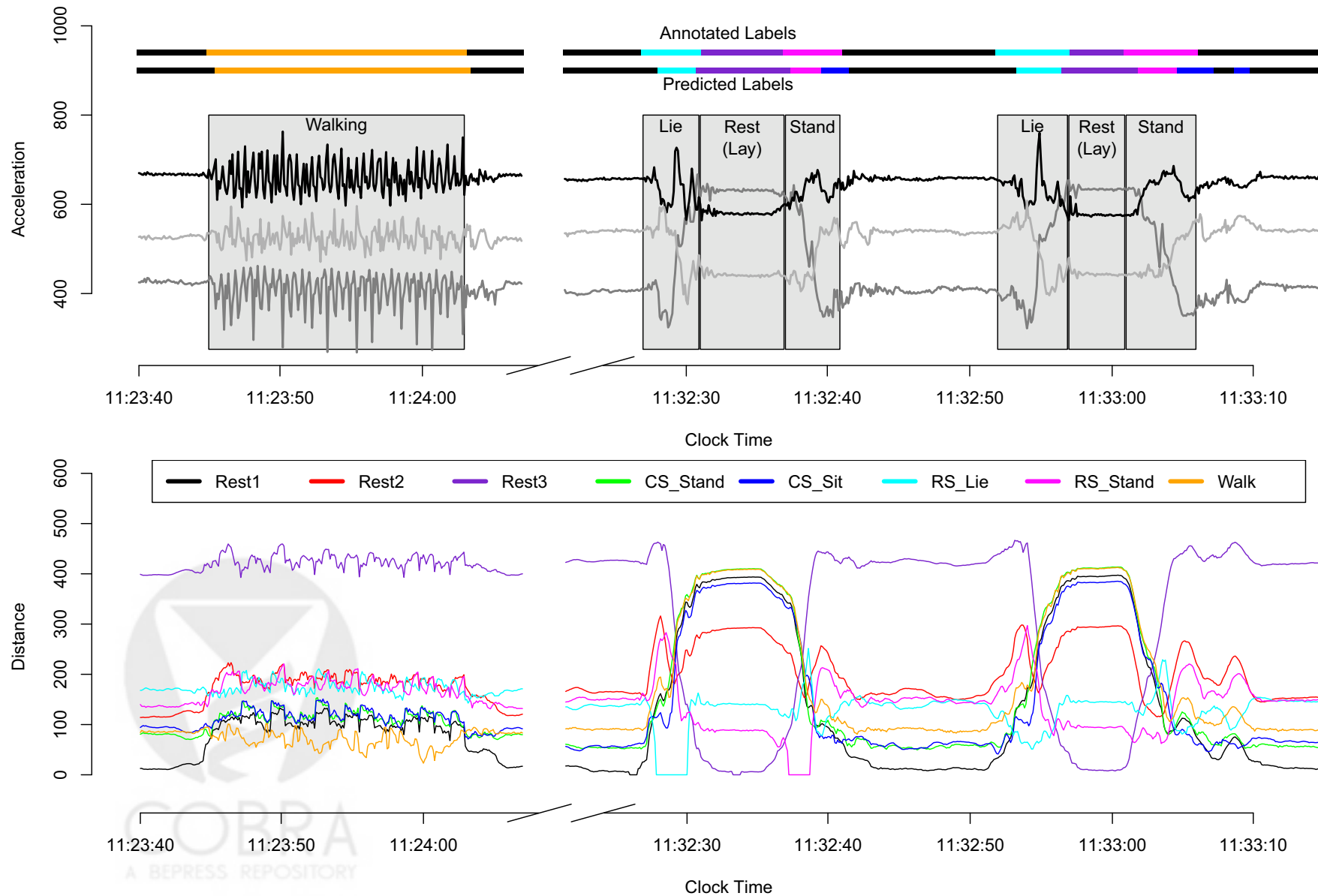
12

Figure 4: Observer-defined annotations and predictions for two segments of accelerometer data with several activity types. Curves giving the smallest distance between movelets and each chapter are displayed.

For the accelerometer data displayed in Figure 1 (one segment of walking and two replicates of lie-rest-stand), the lower panel of Figure 4 shows the minimum distance between each unlabeled movelet and all movelets contained in the reference chapters as a collection of distance curves. The preliminary labels $L_i^*(t)$ are taken to be the chapter title with smallest distance. Next, the prediction $\hat{L}_i(t)$ is determined via a majority vote in which each element of $\{L_i^*(t), L_i^*(t+1), \ldots, L_i^*(t+H)\}$ ($t \in [0, T_i - H]$) is considered a single vote. At the top of Figure 4 are the observer-annotated (top colored bar) and predicted labels (bottom color bar) that accompany the accelerometer data. A comparison of the annotations and predictions indicates generally high agreement between these time series. In particular, there is broad overlap between the prediction and annotation of walking and resting periods as well as the location of the shorter activities lying and standing. Moreover, there is generally reasonable separation between the distance curve corresponding the the correct chapter and the remaining chapters, indicating the ability of the movelet-based analysis to distinguish between activity types. In two regions, the distance curves are zero – these depict the first replicate of the "Lie from Stand" and "Stand from Lie" activities, and were used to construct their respective activity chapters. Isolated misclassifications in the preliminary labels, such as those that take place in the middle of walking period, are in effect smoothed by the majority-voting step which prevents single activity labels from disagreeing with its neighbors.

On the other hand, as shown in the right segment of Figure 4, the annotated labels for the shorter activities have much longer time durations than the predicted intervals. This is most likely due to a combination of early and late stop points in the annotations and time spent transitioning between activities. For example, when a subject is asked to sit from a standing position, there is a brief pause as the new movement is begun; similarly, when rising to a standing position, there is a short period of stabilization as the movement is completed. The "true activity" at these time points is not clearly defined, but the annotations are seen to be conservative in starting and stopping short activities, whereas the predictions extend neighboring (well-predicted) resting periods. This contrast can negatively affect the apparent prediction accuracy, although many of the activities are correctly identified.

Let V_i^a be the amount of time spent performing activity a (measured by $L_i^{obs}(t)$) and $\hat{V}_i^a$ be the

predicted amount of time spent performing activity a. For each subject and visit, in Table 3, we report $\hat{V}_i^{a'}/V_i^a$ for all activities a, a'.

Table 3 reinforces the observations from Figure 4 that long continuous activities, like resting and walking, are better predicted than short activities, like standing from a chair. In fact, with the exception of subject 1 at visit 1, all resting states are accurately predicted more that 99% of the time, and walking is accurately predicted between 68% and 80% of the time. However, short activities seem to be fairly poorly predicted, and are often mistaken for one of the resting states. Again, this apparent shortcoming stems from two major factors: $i.$ these activities are undertaken for very short periods, so even minor misclassification can greatly impact results, and more importantly $ii.$ the observer-provided annotations for these short activities are inaccurate.

3.3   Refined Results

A comparison of our initial predictions, the observer defined annotations and the raw accelerometer data indicate that a gold standard for $L_i(t)$, the true activity labels associated with acceleration data, is not given by $L_i^{obs}(t)$, the observer's annotations. Thus, we next create a "combined observer" to define activity labels $L_i^{com}(t)$ by synthesizing all available information. Primarily, this resulted in designating times between two distinct activities as "transition times", rather than misleadingly assigning these periods to one or the other activity. The new activity labels are shown in Figure 5, and a comparison of labels $L_i^{com}(t)$ and predictions $\hat{L}_i(t)$ is given in Table 4. These demonstrate the large improvements in prediction accuracy that arise from improvements in the standard used to define true activity labels. We contend that these findings indicate that: 1) accurate labeling is crucial to prediction algorithm training; 2) a large source of prediction inaccuracies can reliably be traced to human labeling; and 3) prediction accuracy results reported in the literature are hard to compare because data use different labeling protocols.

The construction of the combined observer also illustrates the feedback from the movelet-based prediction algorithm to the annotations. Periods that were largely misclassified using $L_i^{obs}(t)$ as a reference, and that were labeled as "transitions" in $L_i^{com}(t)$, are periods where the distance between an unlabeled movelet and those in the reference dictionary is large. Thus,

15

Subject 1 Visit 1

	Prediction							
Truth	Rest1	Rest2	Rest3	CS_Stand	CS_Sit	RS_Lie	RS_Stand	Walk
Rest1	78.3%	0	0	21.7%	0	0	0	0
Rest2	0	100%	0	0	0	0	0	0
Rest3	0	0	100%	0	0	0	0	0
CS_Stand	4.3%	30.5%	0	55.3%	0	0	6.4%	3.5%
CS_Sit	11.9%	37.3%	0	23.1%	27.6%	0	0	0
RS_Lie	0	0	22.9%	49.4%	0	27.7%	0	0
RS_Stand	0	0	21.3%	59.8%	0	0	18.9%	0
Walk	9.6%	0	0	10.4%	0.5%	0	0	79.5%

Subject 1 Visit 2

	Prediction							
Truth	Rest1	Rest2	Rest3	CS_Stand	CS_Sit	RS_Lie	RS_Stand	Walk
Rest1	96.8%	0	0	0	2.7%	0	0	0.4%
Rest2	0	99.9%	0	0	0.1%	0	0	0
Rest3	0	0	100%	0	0	0	0	0
CS_Stand	25.6%	25.6%	0	40.2%	8.5%	0	0	0
CS_Sit	25.7%	12.8%	0	0	57.8%	3.7%	0	0
RS_Lie	40.3%	0	14.9%	0	0	44.8%	0	0
RS_Stand	0	0	21.1%	2.8%	39.4%	0	36.6%	0
Walk	18.8%	0	0	0.2%	0.6%	1.1%	0	79.3%

Subject 2 Visit 1

	Prediction							
Truth	Rest1	Rest2	Rest3	CS_Stand	CS_Sit	RS_Lie	RS_Stand	Walk
Rest1	99.9%	0	0	0	0.1%	0	0	0
Rest2	0.7%	99.2%	0	0.1%	0.1%	0	0	0
Rest3	0	0	100%	0	0	0	0	0
CS_Stand	10.9%	16.4%	0	57.3%	0	0	10.0%	5.5%
CS_Sit	11.1%	44.4%	0	4.6%	34.6%	5.2%	0	0
RS_Lie	10.6%	0	50.6%	0	0	27.1%	11.8%	0
RS_Stand	36.8%	0	24.6%	0	0	0	38.6%	0
Walk	22.1%	0	0	0.3%	0	0.2%	1.0%	76.4%

Subject 2 Visit 2

	Prediction							
Truth	Rest1	Rest2	Rest3	CS_Stand	CS_Sit	RS_Lie	RS_Stand	Walk
Rest1	100%	0	0	0	0	0	0	0
Rest2	0	100%	0	0	0	0	0	0
Rest3	0	0	100%	0	0	0	0	0
CS_Stand	7.9%	40.4%	0	46.1%	0	0	0	5.6%
CS_Sit	33.3%	20.6%	0	9.8%	35.3%	0	0	1.0%
RS_Lie	42.6%	0	31.5%	0	0	25.9%	0	0
RS_Stand	34.4%	0	34.4%	21.3%	0	0	9.8%	0
Walk	31.3%	0	0	0.1%	0	0	0	68.6%

Table 3: Comparison of observer-annotated labels $L_i^{obs}(s)$ and the predicted labels $\hat{L}_i(s)$, expressed as the proportion of the predicted time spent engaged in an activity and the time spent engaged in the activity according to the annotated activity labels.

Subject 1 Visit 2

Truth	Prediction							
	Rest1	Rest2	Rest3	CS_Stand	CS_Sit	RS_Lie	RS_Stand	Walk
Rest1	98.8%	0	0	0%	1.2%	0	0	0
Rest2	0	100%	0	0	0	0	0	0
Rest3	0	0	100%	0	0	0	0	0
CS_Stand	7.5%	0%	0	75.0%	17.5%	0	0%	0%
CS_Sit	4.2%	1.4%	0	0%	88.8%	5.6%	0	0
RS_Lie	0	0	16.7%	0%	0	83.3%	0	0
RS_Stand	0	0	0%	8.0%	0	0	92.0%	0
Walk	13.2%	0	0	0%	0.4%	1.1%	0	85.3%

Table 4: Table of prediction agreement for Subject 1 Visit 2, using the combined observer
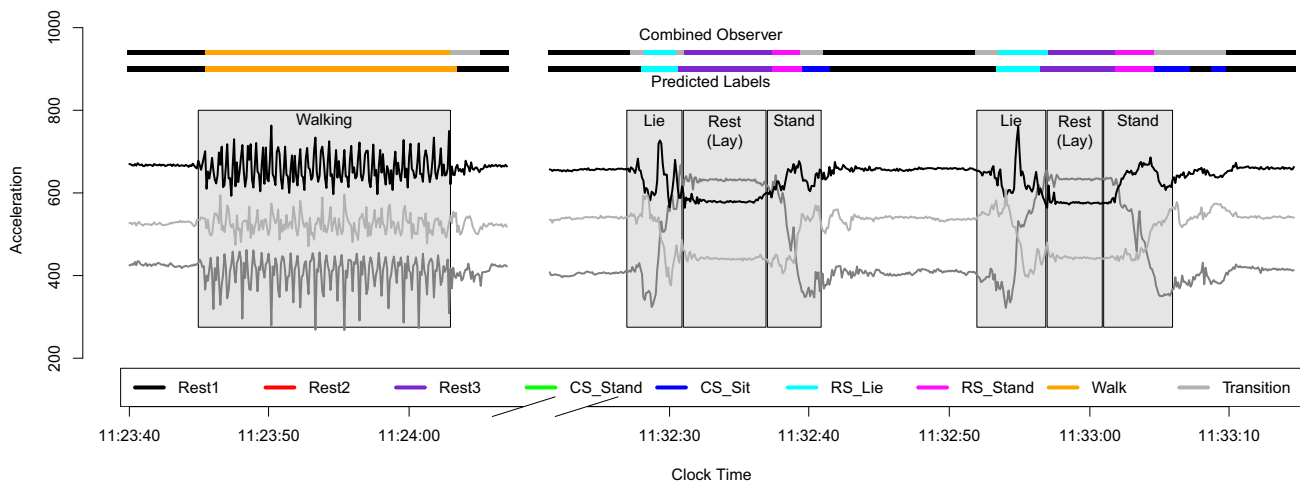

Figure 5: Comparison of "combined observer" annotations, based on observed-defined annotations and an inspection of the raw accelerometer data, and predicted labels.

movelets that don't match well to any known reference can be quickly identified. In observational studies, this facilitates the recognition of movements that are not included in any dictionary or are otherwise abnormal.

4 Discussion

Understanding physical activity is a key component in public health studies of subject function. However, standard measures of physical function such as activities of daily living questionnaires are subject to substantial measurement error. Emerging accelerometer technologies allow

the collection of real-time, real-world activity data and may alleviate many of the issues with retrospective self-report data collection.

In this paper we propose a method for activity classification built around the "movelet" as a basic element of movements. Using movelets with known activities, we construct reference chapters and dictionaries; given an unlabeled movelet, we find its closest match in the reference and use the match's label as a basis for prediction. Thus, our method is built on the intuition that movements with similar component acceleration patterns are likely to be generated by the same activity. This allows the method, and the matches it provides, to be quickly evaluated based on visual inspection of the accelerometer time series. Moreover, the extension to large datasets in which subjects are observed for hours or days is direct, because activity prediction is local in time. Finally, our method accurately predicts short activities, such as taking a few steps, as well as relatively rare and low-frequency movements such as rising from a chair.

Several directions exist for improving the movelet-based method. Focusing on the predictions for a single subject, transition models could naturally encode information about the order of movements and the likelihood of switching between them. Similarly, smoothing the distance functions (shown in Figures 4 and 5) would allow neighboring time points to influence the prediction at the current time. Augmenting dictionaries to include objects other than movelets, for instance by adding measures of mean and variation, could improve predictions. Our method can also be extended to increase the understanding of heterogeneity in acceleration patterns between and within subjects. For instance, constructing a multi-subject dictionary would necessitate an understanding of movement fingerprints across several subjects.

Our results and methods suggest three improvements that could help the deployment of this technology to large epidemiological studies. First, there is an increasing need for developing an accelerometer whose axes are always oriented with respect to gravity. This could probably be done by incorporating a gyroscope. This would resolve the problem of interpretation of the accelerometry data, especially in realistic scenarios where people wear these devices for extended periods of time. Second, the study could be more accurate if a human observer goes to the home of the participants, explains the use of the device, helps setting it up and conducts a short testing period using a known sequence of common activities whose duration and type

is carefully annotated. This would resolve the problem of subject-specific training of prediction algorithms in the home environment and not in a lab. It would also place a smaller burden on the participants. Finally, replication and calibration pre-studies should be conducted to ensure that prediction algorithms perform well on new subject or visit data.

References

A. A. Atienza and A. C. King. Comparing self-reported versus objectively measured physical activity behavior: A preliminary investigation of older filipino american women. *Research quarterly for exercise and sport*, 76(3):358–362, 2005.

J. Bai. Accelerometer-based prediction of activity for epidemiological research. Master's thesis, Johns Hopkins University, 2011.

L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In *Proceedings of the 2nd International Conference on Pervasive Computing*, pages 1–17. Springer, 2004.

J. Boyle, T. Karunanithi, T. Wark, W. Chan, and C. Colavitti. Quantifying functional mobility progress for chronic disease management. In *28th Annul Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5916–5919, New York, 2006.

J. B. Bussmann, W. L. Martens, J. H. Tulen, F. C. Schasfoort, H. J. van den Berg-Emons, and H. J. Stam. Measuring daily behavior using ambulatory accelerometry: the activity monitor. *Behavior Research Methods, Instruments, & Computers*, 33(3):349–356, 2001.

M. Ermes, J. Pärkka, J. Mäntyjärvi, and I. Korhonen. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE Transactions on Information Technology in Biomedicine*, 12:20–26, 2008.

A. R. Feinstein, B. R. Josephy, and C. K. Wells. Scientific and clinical problems in indexes of functional disability. *Annals of Internal Medicine*, 105:413–420, 1986.

P. S. Freedson, E. Melanson, and J. Sirard. Calibration of the computer science and applications, inc. accelerometer. *Medicine & Science in Sports & Exercise*, 30(5):777–781, 1998.

P. M. Grant, C. G. Ryan, W. W. Tigbe, and M. H. Granat. The validation of a novel activity monitor in the measurement of posture and motion during everyday activities. *British Journal of Sports Medicine*, 40:992–997, 2006.

P. M. Grant, P. M. Dall, S. L. Mitchell, and M. H. Granat. Activity-monitor accuracy in measuring step number and cadence in community-dwelling older adults. *Journal of Aging and Physical Activity*, 16:204–214, 2008.

D. Hendelman, K. Miller, C. Baggett, E. Debold, and P. Freedson. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Medicine & Science in Sports & Exercise*, 32:442–449, 2000.

K. Kiani, C. J. Snijders, and E. S. Gelsema. Computerized analysis of daily life motor activity for ambulatory monitoring. *International Journal of Technology Assessment in Health Care*, 5:307–318, 1997.

K. Kiani, C. J. Snijders, and E. S. Gelsema. Recognition of daily life motor activity calsses using an artificial neural network. *Archives of Physical Medicine and Rehabilitation*, 79:147–154, 1998.

A. Krause, D. P. Sieiorek, A. Smailagic, and J. Farringdon. Unsupervised, dynamic identification of physiological and activity context in wearable computing. In *Proceedings of the 7th International Symposiu on Wearable Computers (White Plains, NY)*, pages 88–97. IEEE Computer Society, 2003.

J. Mantyjarvi, J. Himberg, and T. Seppanen. Recognizing human motion with multiple acceleration sensors. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, pages 747–752. IEEE Press, 2001.

I. McDowell and C. Newell. *Measuring Health: A Guide to Rating Scales and Questionnaires*. Oxford University Press, New York, 1987. ISBN 0-19-516567-5.

A. Nguyen, D. Moore, and I. McCowan. Unsupervised clustering of free-living human activities using ambulatory accelerometry. *29th Annual Conference of the IEEE Engineering in Medicine and Biology Society (Lyon)*, pages 4895–4898, 2007.

J. Pärkka, M. Ermes, P. Korpipää, J. Mäntyjärvi, J. Peltola, and Ilkka Korhonen. Activity classification using realistic data from wearable sensors. *IEEE Transactions on Information Technology in Biomedicine*, 10:119–128, 2006.

R. R. Pate, M. Pratt, S. N. Blair, W. L. Haskell, C. A. Macera, C. Bouchard, D. Buchner, W. Ettinger, G. W. Heath, A. C. King, A. Kriska, A. S. Leon, B. H. Marcus, J. Morris, R. S. Paffenbarger, K. Patrick, M. L. Pollock, J. M. Rippe, J. Sallis, and J. H. Wilmore. Physical activity and public health. a recommendation from the centers for disease control and prevention and the american college of sports medicine. *Journal of the American Medical Association*, 273:402–407, 1995.

J. W. Picone. Signal modeling techniques in speech recognition. In *Proceedings of the IEEE*, volume 81(9), pages 1541–1546, 2005.

D. M. Pober, J. Staudenmayer, C. Raphael, and P. S. Freedson. Development of novel techniques to classify physical activity mode using accelerometers. *Medicine & Science in Sports & Exercise*, 38(9):1626–1634, 2006.

N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman. Activity recognition from accelerometer data. In *Proceedings of the Seventeenth Conference on Innovative Applications of Artificial Intelligence*, pages 1541–1546. AAAI Press, 2005.

J. Staudenmayer, D. Pober, S. Crouter, D. Bassett, and P. Freedson. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *Journal of Applied Physiology*, 107(4):338–345, 2009.

R. P. Troiano, D. Berrigan, K. W. Dodd, L. C. Mâsse, T. Tilert, and M. McDowell. Physical activity in the united states measured by accelerometer. *Medicine & Science in Sports & Exercise*, 40(1): 181–188, 2008.

G. J. Welk, S. N. Blair, S. Jones, and R. W. Thompson. A comparative evaluation of three accelerometry-based physical activity monitors. *Medicine & Science in Sports & Exercise*, 32: 489–497, 2000.

K. Zhang, F. X. Pi-Sunyer, and C. N. Boozer. Measurement of human daily physical activity. *Obesity Research*, 11:33–40, 2003.

K. Zhang, F. X. Pi-Sunyer, and C. N. Boozer. Improving energy expenditure estimation for physical activity. *Medicine & Science in Sports & Exercise*, 36:883–889, 2004.