Real-Time Recognition of Physical Activities and Their Intensities Using Wireless Accelerometers and a Heart Rate Monitor

Emmanuel Munguia Tapia[1], Stephen S. Intille[1],
William Haskell[3], Kent Larson[1], Julie Wright[2], Abby King[3], and Robert Friedman[2]
[1]*House_n, Massachusetts Institute of Technology, Cambridge, MA, USA*
[2]*Boston Medical Center, Boston, MA USA;* [3]*Stanford Medical School, Palo Alto, CA, USA*
emunguia | intille @mit.edu

Abstract

In this paper, we present a real-time algorithm for automatic recognition of not only physical activities, but also, in some cases, their intensities, using five triaxial wireless accelerometers and a wireless heart rate monitor. The algorithm has been evaluated using datasets consisting of 30 physical gymnasium activities collected from a total of 21 people at two different labs. On these activities, we have obtained a recognition accuracy performance of 94.6% using subject-dependent training and 56.3% using subject-independent training. The addition of heart rate data improves subject-dependent recognition accuracy only by 1.2% and subject-independent recognition only by 2.1%. When recognizing activity type without differentiating intensity levels, we obtain a subject-independent performance of 80.6%. We discuss why heart rate data has such little discriminatory power.

1. Introduction

Automatic detection of physical activity (PA) might enable new types of health assessment and intervention tools that help people maintain their energy balance and stay physically fit and healthy. Recent research has shown that wearable accelerometers can be used to reliably detect some physical activity types when tested on small datasets (e.g.[1-4]). We are unaware; however, of work showing the same algorithm can detect not only activity type but also, in some cases, the same activity at different intensities. Furthermore, most work with accelerometers has either used cumbersome wired sensors [3] or sensors that store data locally for off-line processing [1, 4, 5]. Here we show how wireless sensors transmitting raw data in real-time (and thus susceptible to signal loss) could be used for automatic PA and PA-intensity recognition.

Past work in recognizing activities from accelerometer data has used computationally intensive supervised classification algorithms that typically require offline analysis [2, 3, 6]. In this work, we utilize fast Decision Tree (DT) classifiers (as in [1, 4, 5]) with a set of efficiently computable features to achieve real-time performance on current PCs. DTs can overfit to the data if sufficiently diverse training sets are not used. Therefore, we test on a relatively large dataset (compared with prior work) consisting of 30 gymnasium activities (see Table 1) collected from 21 participants by two different teams.

We also study the usefulness of heart rate (HR) data in discriminating the intensity of activities. HR may be useful since it correlates with energy expenditure for aerobic exercise; however, alone it provides little information about activity type, and it is influenced by other factors such as emotional states, ambient temperature, and fitness level. HR also responds and stabilizes slowly. In this work we explore the "best-case" scenario of how well activity type and, for some activities, intensity can be recognized when we place one accelerometer at each limb and the hip, and a HR monitor on the chest.

2. System overview and data collection

The PA recognition system consists of five triaxial wireless wearable accelerometers sensors sampling at 30Hz, a wireless HR monitor based on the Polar chest strap (Wearlink), and a laptop computer with a wireless receiver. These sensors allow data to be collected from multiple body points simultaneously without constraining movement [7] .

To acquire training data for the PA recognition system, a total of 21 participants between 18 and 65 years old and with varying levels of physical fitness were recruited at two separate medical labs: (1) The Boston Medical Center and (2) Stanford Medical School. Using cotton elastic sweat bands or non-restrictive adhesive bandages, researchers placed the accelerometers on each subject, with one at each of the following locations: top of the dominant wrist just behind the wrist joint, side of the dominant ankle just above the ankle joint, outside part of the dominant upper arm just below the shoulder joint, on the upper part of the dominant thigh, and on the dominant hip, as

1-4244-1453-9/07/$25.00 ©2007 IEEE.

indicated in Figure 1b. All the accelerometers were ±10G except the accelerometer on the hip, which was ±2G. The HR monitor was worn on the chest.

After the sensors were placed, each participant was asked to sit still and, after a stabilization period, resting HR was measured by measuring pulse for one minute. The participant's age-predicted maximum HR (MHR=220-age) was calculated. A combined 21 participants each performed 30 gymnasium activities for 2min each, with 12 and 9 datasets being collected from each site, respectively. The list of gymnasium activities, broken down by type and intensity differences, is shown in Table 1.

The activities with different intensity levels are *walking*, *cycling*, and *rowing*. For *walking*, we varied intensity by changing the treadmill speed (e.g. 2, 3, and 4 mph) and inclination (e.g. 4, 8, and 12 degrees). For *cycling*, we varied the cycle speed (e.g. 60, 80, and 100 rpm) and the cycle resistance level to settings that participants subjectively considered equivalent to light, moderate, and hard. Finally, for *rowing*, we kept the rowing speed constant at 30 spm and varied the resistance until reaching levels that participants considered light, moderate, and hard.

In total, 16.6 hours of usable annotated data were collected. In the remainder of this work, G1 refers to the gym dataset collected from site one and G2 refers to the dataset collected from site two. Dataset G2 differs slightly from dataset G1; due to lab constraints, data for the *move weight*, and *calisthenics (cali.)* activities were not collected, and the *rowing* activity was substituted by *arm ergometry*. Nevertheless, dataset G2 contains the same number of activities with different intensity levels as G1 (*walking*, *cycling*, and *arm ergometry*). Researchers interested in using these datasets should contact the authors.

3. Activity recognition algorithm

In the training step, data segments not labeled as one of the target activities listed in Table 1 are discarded. The 15 acceleration data streams (x, y, and z axes) were then broken into 50% overlapping sliding windows of length 4.2s and independently interpolated using cubic spline interpolation to fill out missing sensor values lost during wireless transmission. In Section 4, we describe how 4.2s was selected as the window size. If the percentage of samples lost inside a given window of length 4.2s was greater than 20% for any of the accelerometer axis streams, the window was discarded. To smooth out the noisy HR data, we applied a running average filter over the past 30s of data. The HR data is then segmented by accumulating the data over 30s windows from the end time of each acceleration

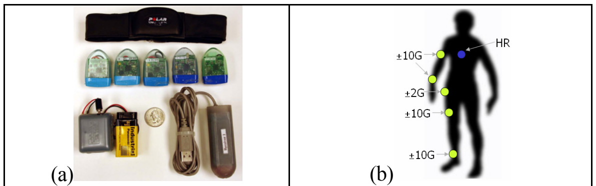

Figure 1. (a) Five 3-axis wireless accelerometers, a heart rate monitor, and USB wireless receiver, and (b) Placement of the sensors on the body.

Type	Intensity	Type	Intensity
Lying down	N/A	Cycling	Moderate at 80 rpm
Standing	N/A	Cycling	Light at 80 rpm
Sitting	N/A	Cycling	Light at 60 rpm
Sitting	Fidget feet legs	Cycling	Light at 100 rpm
Sitting	Fidget hands arms	Rowing	Light at 30 spm
Walking	2mph 0% grade	Rowing	Hard at 30 spm
Walking	3mph 0% grade	Rowing	Moderate 30 spm
Walking	3mph at 4% grade	Carry weight 2mph	N/A
Walking	3mph at 8% grade	Move weight high	N/A
Walking	3mph at 12% grade	Move weight low	N/A
Walking	4mph at 0% grade	Move weight side	N/A
Running	5mph at 0% grade	Cali. Bicep curls	N/A
Ascend stairs	N/A	Cali. Jumping jacks	N/A
Descend stairs	N/A	Cali. Push ups	N/A
Cycling	Hard at 80 rpm	Cali. Sit ups	N/A

Table 1. Activities studied in this work.

window going backwards in time. HR windows are discarded when no samples are available for a given window. Overall, only 3.2% of the data (32.2m of 16.6h) collected were discarded due to accelerometer signal loss and 1.9% due to HR signal loss.

Time domain and frequency domain features are then computed for each 4.2s window. Here we use the area under curve (AUC) and variance to capture signal variability, mean distances between axes and mean to capture sensor orientation with respect to ground for postures, entropy to differentiate activity type, correlation coefficients to capture simultaneous motion of limbs, and FFT peaks and energy to discriminate between intensities. All the features are computed over each acceleration axis. The only feature computed over the HR data was the number of heart beats above the resting HR value (BPM-RHR). Finally, we used the WEKA toolkit [8] to evaluate the performance of the C4.5 DT [9] (pruned) and the NB classifier using one Gaussian distribution per feature per class.

4. Evaluation

To evaluate the performance of the recognition algorithm, we computed the true positive rate, false positive rate, precision, recall, and F-Measure over the segmented classes using subject-dependent and subject-independent training. In subject-dependent training, we performed 10-fold cross-validation over each subject's data and averaged the results over all the subjects. In subject-independent training, we trained the algorithms with the data of all the subjects but one and tested the performance on the left-out subject. We repeated this procedure for as many subjects as we had and averaged the results. To better understand the performance of the algorithms, the results are clustered into three categories based on activity type: (1) postures (e.g. *lying down*, *standing*, and *sitting*), (2) activities with multiple intensities (*walking*, *rowing/arm ergometry*, and *cycling*), (3) and other activities (*running*, *calisthenics*, *move weight*, and *using stairs*).

4.1. Subject dependent analysis

We first determined the most appropriate window length to use (4.2s) by varying the window length from 0.5 to 17 seconds and measuring the performance of the C4.5 classifier over the datasets. A window of 4.2s is long enough to obtain a good accuracy (74-86%) while minimizing real-time classification delay. Using a similar strategy, we determined that using only two FFT peaks provided good performance. We then performed feature selection over subsets of all the features using the wrapper method and the C4.5 classifier. The most powerful features found, in decreasing order of importance, were the area under curve (93.1% accuracy using only this feature), mean distances between axes (92.1%), mean (91.3%), variance (88.7%), FFT peaks (86.1%), and correlation coefficients (74.8%). Consequently, we measured the performance of the C4.5 DT and the Naïve Bayes (NB) classifier over the best performing subset of these features that we call *variant* features: area under curve (15 values), mean distances between axes (15), mean (15), variance (15), FFT peaks (60), correlation coefficients (105), energy (15), and entropy (15) for a total feature vector with 255 values. Table 2 shows the performance measures using these features over both datasets (4.2s windows).

Table 2 shows that the performance is comparable using both classifiers. As a result, from this point on, we present results only for the C4.5 decision tree classifier. From Table 2, we can also observe that the performance is higher for dataset G2. We believe that this is because G2 contains 8 fewer activities than dataset G1 as explained in Section 2. Another important result is that we have achieved an average false positive rate of only 0.15% over both datasets.

A problem we encountered was that features with the highest discriminant power, such as AUC and the mean, are strongly sensitive to the acceleration signal magnitude and thus dependent on sensor orientation, and calibration. Consequently, we considered utilizing only features invariant to the signal magnitude. After evaluating the performance of subsets of these features using the C4.5 DT classifier over the datasets, we found that the best subset of features was: mean distances between axes, variance, energy, FFT peaks, and correlation coefficients, (225 values in total).

Dataset	**Classifier**	**Activity**			**Total Accuracy (%)**
		Postures (%)	**Other (%)**	**Intensity (%)**	
G1	C4.5	FP: 0.04 P: 98.7 R: 98.6 F: 98.7	FP: 0.20 P: 93.9 R: 93.8 F: 93.8	FP: 0.28 P: 92.2 R: 92.2 F: 92.2	93.7 ± 1.5
G2	C4.5	FP:0.04 P: 98.9 R: 99.1 F: 99.0	FP: 0.19 P: 96.3 R: 96.1 F: 96.2	FP: 0.19 P: 96.0 R: 96.0 F:96.0	96.0 ± 0.9
G1	NB	FP: 0.06 P: 98.1 R: 99.2 F: 98.7	FP: 0.22 P: 93.7 R: 93.6 F: 93.5	FP: 0.30 P: 92.1 R: 92.3 F: 92.1	93.3 ± 2.2
G2	NB	FP: 0.09 P: 97.7 R: 98.5 F: 98.1	FP: 0.05 P: 99.0 R: 95.3 F: 97.1	FP: 0.13 P: 97.2 R: 98.0 F:97.6	97.6 ± 0.8

Table 2. False positives (FP), precision (P), recall (R) and F-measure (F) for subject-dependent analysis using the *variant* features

Dataset	**Features**	**Activity**			**Total Accuracy (%)**
		Postures (%)	**Other (%)**	**Intensity (%)**	
G1	Invariant	FP: 0.06 P: 98.1 R: 98.8 F: 98.4	FP: 0.23 P: 93.5 R: 93.1 F: 93.1	FP: 0.32 P: 91.5 R: 92.0 F: 91.6	92.8 ± 2.5
G2	Invariant	FP: 0.09 P: 97.5 R: 98.4 F: 98.0	FP: 0.05 P: 99.1 R: 94.6 F: 96.8	FP: 0.16 P: 96.6 R: 97.57 F: 97.0	97.1 ± 1.2
G1	Invariant + HR	FP: 0.07 P: 98.0 R: 98.4 F: 98.2	FP: 0.19 P: 93.8 R: 93.7 F:93.8	FP: 0.20 P: 94.3 R: 94.2 F: 94.3	94.8 ± 1.7
G2	Invariant + HR	FP: 0.06 P: 98.3 R: 98.5 F:98.4	FP: 0.29 P: 94.8 R: 95.0 F: 94.9	FP: 0.14 P: 97.1 R: 96.9 F:97.0	96.9 ± 0.7

Table 3. False positives (FP), precision (P), recall (R) and F-measure (F) for subject-dependent analysis using the C4.5 DT and *invariant* features.

Table 3 presents the performance using these features we call invariant. Overall, the C4.5 classifier achieved an average accuracy of 94.9% on both datasets, an accuracy as good as the one obtained using the non-invariant features (94.9%).

After analyzing Table 3 and the confusion matrices, we observed that most of the errors were occurring when the classifier was trying to discriminate between the different *intensity* levels of the same activity. To further explore this, we created a new dataset that we call the *no-intensities* dataset where activities with different intensities, such as all the walking activities, were merged into one class. When we trained the C4.5 classifier using the *invariant* features over this new dataset, we found an improved performance of 97.3 ± 0.7 on G1 and 98.7 ± 0.4 on G2, or an average improvement of 3.3%.

The next step was to investigate if HR data could improve the discrimination among the intensity levels of an activity. To test this, we added the number of heart beats above resting HR (BPM-RHR) to the *invariant* features. Table 2 shows the result. The average performance over both datasets is 95.8%, an improvement of 1.2%.

In order to investigate why the HR feature has such a low impact on improving the discrimination between intensity levels, we trained the C4.5 classifier using only the HR. The recognition performance obtained was 34.0 ± 6.0 for G1 and 49.2 ± 6.7 for G2 using subject-dependent training. Overall, the results are higher for G2 because it contains fewer activities (8) than G1. After plotting misclassification histograms, we observed that the errors were concentrated at the beginning and end of activities. This is because HR lags physical activity and remains altered once the activity has finished (errors at the end of activity or beginning of the next one). Furthermore, for vigorous activities of short duration such as walking up stairs, HR increases constantly, resulting in classifications errors all across the activity.

4.2. Subject independent analysis

For the subject-independent analysis, we repeated the same procedure as the one followed in the previous section. Table 4 shows the results over the invariant features with and without incorporating HR.

The overall performance is relatively low, with an average accuracy of 56.3% (FP: 1.5%) using the C4.5 classifier on both datasets. When the HR feature is added, the average performance improves only 2.1%. To better understand why the HR feature has such a low impact on improving the discrimination between intensity levels, we trained the C4.5 classifier using only the HR feature. The recognition performance is as follows: 12.3 ± 1.7 for G1 and 14.4 ± 2.7 for G2 using subject-independent training. Consequently, we believe that HR does not improve discrimination in subject-independent training because subjects have different fitness levels and the number of beats above resting HR (BPM-RHR) is different for two subjects performing the same activity but with different levels of physical fitness. To minimize the effects of the physical fitness level of each individual, we repeated the experiment when HR (BMP) is normalized to lie between resting HR (RHR) and maximum HR (MHR) for each individual (MHR estimated as 220-age). Using this normalization, two individuals with different fitness level performing the same activity could have different BMP values, but relative to their MHR, they could be performing in the same intensity zone. Unfortunately, the results were similar to those obtained when not scaling the HR data. This may be because the MHR was estimated rather than measured.

After analyzing the confusion matrices we also observed that most of the errors were occurring when the classifier was trying to discriminate between the different intensity levels of an activity. Furthermore, when we train the C4.5 classifier (subject-independent) using the no-intensities dataset and the invariant features, we found an improved performance of 81.1 ± 11.9 for G1 and 80.1 ± 19.4 for G2. This means that (1) the classifier is indeed confusing between intensity levels and (2) that the subject-independent performance when no intensity levels are present is reasonable.

Dataset	Features	Activity			Total Accuracy (%)
		Postures (%)	Other (%)	Intensity (%)	
G1	Invariant	FP: 1.21 P: 66.2 R: 65.0 F: 63.9	FP: 1.02 P: 67.8 R: 68.9 F: 67.8	FP: 2.05 P: 46.4 R: 47.0 F: 46.1	55.6 ± 8.8
G2	Invariant	FP: 0.66 P: 83.7 R: 78.7 F: 80.8	FP: 1.97 P: 61.3 R: 55.5 F: 55.1	FP: 2.55 P: 48.8 R: 53.3 F: 50.3	57.0 ± 13.3
G1	Invariant + HR	FP: 1.15 P: 68.1 R: 67.5 F: 66.5	FP: 1.0 P: 68.8 R: 69.4 F: 68.7	FP: 1.91 P: 49.0 R: 49.0 F: 48.5	58.2 ± 11.0
G2	Invariant + HR	FP: 0.55 P: 86.1 R: 82.3 F: 83.8	FP: 1.99 P: 63.08 R: 46.5 F:51.9	FP: 2.47 P: 50.0 R: 57.6 F: 52.4	58.6 ± 14.9

Table 4. False positives (FP), precision (P), recall (R) and F-measure (F) for subject-independent analysis using the C4.5 DT and *invariant* features.

5. Acknowledgements

Funded by NIH R21 grant CA106745-02. Sensor development supported by NSF grant #0313065.

6. References

[1]. Bao, L. and S.S. Intille, *Activity Recognition from User-Annotated Acceleration Data*, in *Proc. Int'l Conf. Pervasive Comp*. 2004: p. 1-17.

[2]. Lester, J., et al., *A Hybrid Discriminative/Generative Approach for Modeling Human Activities*, in *Proc. Int'l Joint Conf. on Artificial Intelligence*. 2005: p. 766-772.

[3]. Lukowicz, P., et al., *Recognizing Workshop Activity Using Body Worn Microphone and Accelerometers*, in *Proc. Int'l Conf. Pervasive Comp*. 2004: p. 18-32.

[4]. Maurer, U., et al., *Activity Recognition and Monitoring Using Multiple Sensors on Different Body Positions*, in *Proc. Int'l Workshop on Wearable and Implantable Body Sensor Networks*. 2006, p. 113-116.

[5]. Parkka, J., et al., *Activity Classification Using Realistic Data From Wearable Sensors*. IEEE Transactions on inf. Tech. in Biomedicine, 2006. **10**(1): p. 119-128.

[6]. Subramanya, A., et al., *Hierarchical Models for Activity Recognition*, in *Proc. of the IEEE Conference in Multimedia Signal Processing*. 2007: p. 233-237.

[7]. E. Munguia Tapia, et al., *The Design of a Portable Kit of Wireless Sensors for Naturalistic Data Collection*, in *Proc. Int'l Conf. Pervasive Comp*. 2006: p. 117-134.

[8]. Witten, I.H. and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd ed. 2005: Morgan Kaufmann.

[9]. Quinlan, J.R., *C4.5: Programs for machine learning*. 1993, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. 302.