2011 15th Annual International Symposium on Wearable Computers

What can an arm holster worn smart phone do for activity recognition?

Michael Muehlbauer, Gernot Bahle, Paul Lukowicz
Embedded Systems Lab, University of Passau, Passau, Germany
first.lastname@uni-passau.de

Abstract

While mobile phones are increasingly being used in activity recognition, tasks that require arm motion monitoring have so far not been studied on phone platforms. We leverage the fact that upper arm holsters are an increasingly popular way of wearing mobile devices during physical exercise to investigate the suitability of such platforms for arm dominated activity recognition. We focus on (1) user independent recognition from (2) a NULL class dominated, continuous data stream and (3) requiring no special care in device attachment (apart from being placed in a commercial holster). These are 3 key requirements for a realization in a real life mobile "App". We evaluate our methods on a gym exercises data set from 7 users that contains 11'000 individual repetitions of 10 different upper body exercises organized in 700 "sets" (=consecutive repetitions of the same exercise). On set level we achieve a user independent recognition of 93.6%. In over 90% of cases we can also count individual instances with an accuracy of ± 20%.

1. Introduction

The introduction of inertial sensors into smart phones is widely seen as an opportunity for wearable activity recognition to become a main stream technology. Applications ranging from step counters and caloric expenditure monitoring to software characterizing snowboard jumps are commercially available. Building on previous work on multimodal activity recognition (e.g. [10, 12, 4, 6]) researchers have investigated different types of more complex activity recognition using mobile phones (e.g. [13, 7, 9, 3]).

One area widely studied in the wearable community that has so far not benefited from smart phones is gesture related activity recognition (e.g. [5, 1]). The main reason such work does not easily transfer to sensor equipped phones is obvious: it mostly relies on several sensors carefully placed on arms and wrists. However, there are also two subtler but not less important reasons: (1) user dependence and (2) difficulties in spotting the relevant gestures within a large ill defined Null class. Both are particularly grave for arm gesture based systems due to large variability of arm motions.

Thus many gesture oriented activity recognition systems achieve good performance for user specific training only. From the point of view of a smart phone application, where a user would like to buy an App he can use immediately, this is a considerable disadvantage. In addition, many systems assume that the data has been pre-segmented to get rid of such Null class segments using other sources of information (besides arm motions). Using the smart phone as a sole sensor node, such pre-segmentation is often not possible.

Paper Contributions

We address the above issues leveraging the fact that upper arm holsters (see Figure 1) are an increasingly popular way of wearing MP3 players and phones during physical exercise. We demonstrate how such an upper arm mounted system can be used to recognize simple arm based activities on the example of upper body gym exercises. While the recognition of gym exercises with wearable sensors has been demonstrated before (e.g [11, 2]), our work advances the state of the art in three areas. First, we demonstrate a user independent system. Second, we investigate not just the discrimination between exercises, but also the spotting and counting of individual instances in a continuous, diverse, Null class dominated data stream. And third, we rely on a single, upper arm mounted mobile phone as sole sensor, which is much more challenging then more elaborate sensor setups used in previous work. The differences between the exercises are subtle in upper arm signals and the holster based mounting leads to considerable variations in sensor placement and orientation. One limitation of our work is the restriction to upper body exercises involving arm movement; this results from restricting ourselves to only a single phone and no other sensing devices.

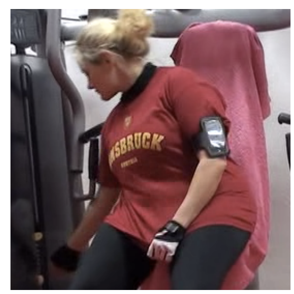

Figure 1. Placement

The paper describes the system concept, the spotting and counting methods and the experimental evaluation. The evaluation has been performed on a data set from 7 users comprising a total of more than 700 individual sets (a set be-

1550-4816/11 $26.00 © 2011 IEEE
DOI 10.1109/ISWC.2011.23

ing a sequence of repetitions of a single exercise at a given machine) with about 11'000 individual exercise instances.

2. Approach

2.1. Problem definition

A typical workout follows a simple pattern: the athlete selects an exercise, performs a number of repetitions, then takes a break. During the break he may drink, stretch, walk around or just rest. Each group of repetitions is called a set. After the break, he repeats the above steps. In total, a typical workout lasts from 60 to 120 minutes and contains up to 10 different exercises and up to 20 sets. Depending on the exercises and workout regimen the athlete follows, each set has between 6 and 40 repetitions.

Keeping the goals defined earlier in mind, our work is motivated by the vision of a user independent "exercise statistics App" requiring nothing more than the device being worn in a holster on the upper arm. Neither user specific training nor more elaborate device attachment efforts should be required. After the workout the App should be able to list the number of repetitions and the timing of all exercises. In particular, all 7 subjects stated that they document sets and repetitions to monitor progress and would like to have that functionality provided automatically.

To that end, we perform three consecutive steps:

1. Segmenting the data into relevant parts (i.e. exercises) and irrelevant noise (i.e. rest, walking, drinking, etc.)
2. Recognizing the exercises performed
3. Counting the repetitions for each set

Before we describe these steps in detail, some information about the hard- and software used: data was recorded with the app ContextLogger on iPhones 3GS and 4. The phones were attached using commercially available holsters from Belkin. Data processing was done with Matlab (segmentation, feature calculation) and University of Waikato's Weka machine learning toolbox (classification).

2.2. Segmentation

As the raw signal consists of large intervals of Null-class (over 80% in our experiments) interspersed with comparatively short exercise periods, most of the signal is noise and has to be eliminated. At the same time, care has to be taken to keep most of the relevant parts. We achieve this by exploiting the periodicity of gym exercises, as they are usually performed with many repetitions. To that end, we first low pass filtered the signal with a cut off frequency of 1/2 Hz; the resulting data was then buffered into large sliding windows (ca. 15 sec window size and 14 sec overlap). For each window, we calculated median crossings (mcr) and the peaks of the autocorrelation (pac) of the signal. High values of both features are very indicative of an exercise being performed. Note that using autocorrelation peak counts grants a certain measure of robustness vs. varying exercise speeds and the weights used. Even if the speed is varied, autocorrelation will still show a peak (albeit a less pronounced one) as long as the basic motion is not too dissimilar. Problems might occur if a subject pauses while in the middle of a set or else if a very high intensity, low repetition count workout routine is followed. From a practical point of view, the latter could, however, be compensated by providing different threshold presets as a choice to the user. The thresholds themselves were determined using a randomly selected, small subsample (5 workouts) of the data, then tested on the remaining workouts. Finally, unifying the sliding windows with a majority decision resulted in a segmentation of the signal in relevant and irrelevant parts.

2.3. Recognizing the exercises

On the remaining relevant data, we calculated a number of features including mean, standard deviation, fluctuation, spectral fluctuation, bandwidth, frequency centroid, 5 cepstrum coefficients and spectral rolloff frequency on sliding windows (1 second window size, 0.5 seconds overlap) on each of the individual axes (Acceleration x y and z and for the Iphone 4 data additionally Gyroscope x y z).

With those features, we then used WEKA for training and testing with a knn-Classifier (k = 3), a C4.5 decision tree and radial kernel SVMs. The knn performed best by $\sim$5%, so the results presented have been obtained using it.

Afterwards, we used the individual segments obtained by the spotting process to perform a modified majority decision on the knn-derived 1 sec frame by frame values: leading and trailing null class elements were judged to be inaccuracies of the spotting process and were cut off; on the remaining core of each segment, a regular majority decision was performed, leading to an exercise set based classification.

Counting the repetitions

Finally, using the segments obtained from the previous step, we performed a peak count on the low pass filtered raw signal to determine how many repetitions of each exercise were performed. Since the peak counts differed from the actual exercise counts both by an offset and a linear factor (some exercises might e.g. have two peaks per repetition instead of one), we calculated linear regression coefficients on a small subsample of the data, then applied those to the raw count results of the entire dataset.

3. Results

3.1. DataSet

The dataset consists of 35 workouts by 7 different persons, 1 of them female. Ages were between 21 to 30 years, experience levels between novice and experienced. The data covers 10 upper body exercises (Butterflies, Chest Press, Latissimus, Abdominal, Upper Back, Shoulder Press, Pulldown, Low Row, Arm Curl, Arm Extension) mixed with a Null class (as users did whatever they wanted during breaks) spanning 83% of the data. This illustrates both the need for a good segmentation algorithm and the sizeable amount of raw data. Recording was usually done in sessions of 1.5 hours, with at most 2 people at once (each wearing a separate device, one person performing their exercises during the breaks of the other subject). On average, the number of different exercises per workout was 9, split into 20 sets, with about 16 repetitions per set. The total number of sets for all 10 exercises was about 700, the total number of repetitions performed more than 11'000. Data was recorded using the iPhone 3GS (21 workouts) and, once available, the iPhone 4 (14 workouts) in a commercial upper arm holster. Users were free to wear the devices as they felt comfortable with no special instructions with respect to fixing or placement.

3.2. Evaluation Strategy

Using the approach we outlined in the previous section, we first tested our segmentation algorithm. Note that it requires no training and is fully user independent. We then apply our recognition method in three different scenarios:

- Training and testing the system on the same person (user dependent mode as baseline)
- Testing on a user on which the system was not trained (user independent classification) also analysing the influence of the number of persons used for training
- User independent classification, adding gyroscope data obtained from the Iphone 4

Finally, we present the results of the set repetition counts.

3.3. Segmentation

Utilizing the techniques described previously, on average we managed to obtain a recall of 95.1% for the relevant data (54.8% precision). At the same time, 82.8% of the null class was filtered out (99.1% precision). Overall, the accuracy (defined as (#true positives + #true negatives) / #of samples) was 85.1%. Table 1 gives a concise overview of the results.

	Rec 1	Prec 1	Rec 0	Prec 0	Accuracy
worst case	0.627	0.156	0.344	0.935	0.496
best case	1.000	0.824	0.951	1.000	0.956
avg. case	0.951	0.548	0.828	0.991	0.851

Table 1. spotting results (1: relevant data, 0: NULL class)

3.4. Person specific classification

As a baseline, we first evaluated the workouts person specifically, testing each workout of every person against the classifier trained by his or her remaining sessions. Not unexpectedly, the frame by frame results were quite encouraging, reaching as high as 94%. Five of the seven results were above 80%. The two subjects that only reached about 70% suffered from the fact that they only contributed 2 and 3 workouts to the dataset, so person specific data was sparse for them. On average, 83.4% of frames were classified correctly. It is worth pointing out that the average of the 5 datasets not suffering from little data would be 89.4%.

Moving from the frame by frame results to the set based recognition boosted performance significantly, the average rising to 93.9%. Even persons P3 and P6, which had shown the worst performance frame by frame achieved about 90% recognition rates. At 99%, P4 achieved an almost perfect result. Figure 2 illustrates these results.

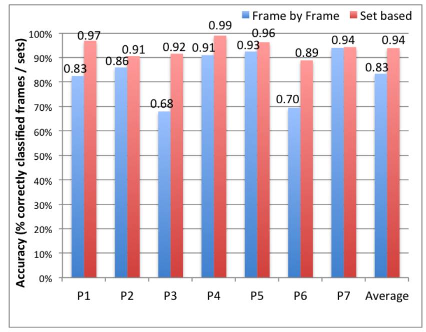

Figure 2. Person specific results

3.5. Person independent classification

Proceeding to the main purposes of our system, person independent recognition, we tested our approach with different numbers of persons as a training group. We evaluated training the classifier on data of 1, 2, 3, 4, 5 and 6 persons, testing with the remaining workouts. As an example, for the 6 person case, this means using subjects 1 to 6 for training, testing with all workouts of subject 7, then training with subjects 1 to 5 and 7 and testing with subject 6 and so on. This amounts to 35 test workouts in the 6 person case and far more in cases with less persons for training. The values presented are the averages of all of these tests. As can be expected, performance starts of very low, at 65.0% of frames correctly recognized for one person (std deviation of 16.1%). Using two persons boosts this by 9%. Adding more subjects still increases gains, but by smaller amounts. Frame by frame recognition performance increases to 87.5% (std deviation of 4.5%) for the 6 person case. While this value is higher than the average of the person specific case (83.4%), it is lower than the average when leaving out the 2 subjects who contributed only a few workouts (89.4%). Figure 3 implies that adding additional persons will not net significant additional gains.

Moving on to recognition of actual training sets once again improves performance significantly by 12.7% for the 1 person case, resulting in a total of 77.7% of correctly classified sets. Training with 6 persons achieves 93.6% recognition of the sets performed; applied to a typical workout, this means that our system misses just one set of repetitions.

3.6. Adding the gyroscope

As mentioned above, the iPhone 4 becoming available during this work provided the opportunity to switch platforms to a phone also providing gyroscope information. Unfortunately, only 14 of the 35 datasets were recorded with the new device, so it was not possible to perform the

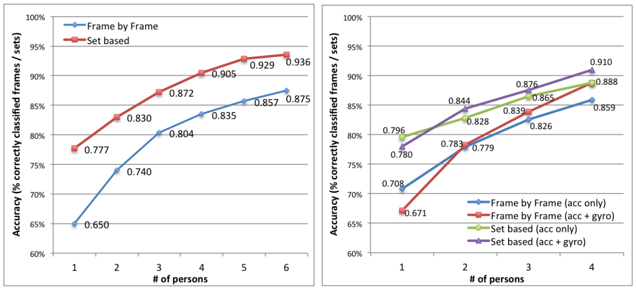

Figure 3. Person independent classification: acceleration and acceleration + gyroscope

same exhaustive analysis as for the acceleration only case. We managed, however, to test the influence of the additional gyroscope information on person independent recognition for training data comprised of 1 to 4 subjects. Training on only 1 person, including gyroscope information into the data actually decreases performance, both for the frame by frame as well as the set based evaluation. This is not unexpected, as the additional data actually increases person specificity. Moving to larger training samples of 2, 3 and 4 persons sees the combined gyroscope and acceleration data outperforming acceleration only. In the 4 person case, frame by frame results without gyroscope are 85.9%; gyroscope information boosts this to 88.8%. In the set based evaluation, acceleration alone reaches an accuracy of 88.8%, while acceleration and gyroscope combined yield 91% recognition. The results of this evaluation are visualized in Figure 3. Since the amount of data is far smaller than in the previous evaluation, it is not possible to compare results one on one. Our results hint, however, that adding a gyroscope is beneficial, though the effect does not seem very pronounced. It might be possible to elicit further gains by applying techniques similar to those described in [8], not using both gyroscope and acceleration at once but selecting the more suitable data source depending on the movement. Also, we would like to point out that many of the exercise trajectories do not contain pronounced rotations of the upper arm, further limiting the usefulness of a gyroscope.

3.7. Counting repetitions

The performance of the repetition counting depends almost entirely on the quality of the segmentation algorithm. While results are slightly modified by the recognition part cutting off leading / trailing null class samples, better defining segments, the influence on the final counts is almost negligible. Thus, we only present the evaluation of the repetition counts for the 6 subjects person independent case.

As Figure 4 shows, the repetitions of 42% of all sets are counted exactly right. 79% of all sets only deviate up to 10% from the correct count, while 92% fall within 20% deviation. At an average number of repetitions per set of about 16, this means that almost all counts are within 3 of the correct number, while about 68% only deviate by 1.

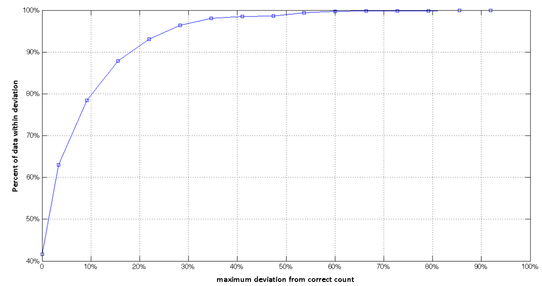

Figure 4. Cumulated repetition counts

4. Conclusion

Clearly, the recognition of upper body gym exercises (or similar activities) with wearable sensors is in itself not a novel result. Instead, the significance lies in showing, on a large, realistic data set, that such a recognition is possible under conditions that one would expect when implementing a real life mobile "App": (1) fully user independent recognition (which means that the system can be pre-trained), (2) the ability to spot the relevant activities despite large, variable Null class and (3) the phone as sole sensing devices with no special attachment requirements (except a commercial arm holster). We believe that our results represent a step towards wider spread of activity recognition applications and more real life impact of wearable computing.[1]

References

[1] E. Berlin, J. Liu, K. van Laerhoven, and B. Schiele. Coming to grips with the objects we grasp: detecting interactions with efficient wrist-worn sensors. In *Proc. of the 4th int. conf. on tangible, embedded, and embodied interaction*, TEI '10, pages 57–64, NY, USA, 2010. ACM.

[2] K.-h. Chang, M. Chen, and J. Canny. Tracking free-weight exercises. In J. Krumm, G. Abowd, A. Seneviratne, and T. Strang, editors, *UbiComp 2007: Ubiquitous Computing*, volume 4717 of *Lecture Notes in Computer Science*, pages 19–37. Springer Berlin / Heidelberg, 2007.

[3] R. K. Ganti, S. Srinivasan, and A. Gacic. Multisensor fusion in smartphones for lifestyle monitoring. In *BSN '10: Proc. of the 2010 Int. Conf. on Body Sensor Networks*, pages 36–43, Washington, DC, USA, 2010. IEEE.

[4] T. Huynh, U. Blanke, and B. Schiele. Scalable recognition of daily activities with wearable sensors. In J. Hightower, B. Schiele, and T. Strang, editors, *Location- and Context-Awareness*, volume 4718 of *Lecture Notes in CS*, pages 50–67. Springer Berlin / Heidelberg, 2007.

[5] H. Junker, O. Amft, P. Lukowicz, and G. Tröster. Gesture spotting with body-worn inertial sensors to detect user act. *Pattern Recogn.*, 41:2010–2024, 2008.

[6] N. Kern, B. Schiele, and A. Schmidt. Multi-sensor activity context detection for wearable computing. In *Ambient Intelligence*, volume 2875 of *Lecture Notes in CS*, pages 220–232. Springer Berlin / Heidelberg, 2003.

[7] A. Khan, Y.-K. Lee, S. Lee, and T.-S. Kim. Human activity recognition via an accelerometer-enabled-smartphone using kernel discriminant analysis. In *Future Inf. Tech. (FutureTech), 2010 5th Int. Conf. on*, pages 1–6, May 2010.

[8] K. Kunze and P. Lukowicz. Dealing with sensor displacement in motion-based onbody activity recognition systems. In *Proc. of the 10th int. conf. on Ubiquitous computing*, UbiComp '08, pages 20–29, NY, USA, 2008. ACM.

[9] J. R. Kwapisz, G. M. Weiss, and S. A. Moore. Activity recognition using cell phone accelerometers. *Human Factors*, page 1018, 2010.

[10] J. Lester, T. Choudhury, and G. Borriello. A practical approach to recognizing physical activities. In *Proc. of Pervasive*, pages 1–16, 2006.

[11] C. Mattmann, O. Amft, H. Harms, G. Troster, and F. Clemens. Recognizing upper body postures using textile strain sensors. In *Wearable Computers, 2007 11th IEEE Int. Symp. on*, pages 29 –36, Oct. 2007.

[12] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher. Activity recognition and monitoring using multiple sensors on different body positions. In *BSN '06: Pro. of the Int. Workshop on Wearable and Implantable Body Sensor Networks*, pages 113–116, Washington, DC, USA, 2006. IEEE.

[13] T. S. Saponas, J. Lester, J. Froehlich, J. Fogarty, and J. Landay. ilearn on the iphone: Real-time human activity classification on commodity mobile phones. Cse technical report, University of Washington, 2008.

[1] This work has been sponsored by the EU ALLOW project