Dataset: Physical Activity Monitoring

Note: This dataset is freely available for academic research, there are no (legal or other) constraints on using the data for scientific purposes. If you want to refer to this dataset in a publication, please use the two references below. If you have any questions or suggestions, please contact Attila Reiss ([firstname].[lastname]@dfki.de). Also, please let us know if you have any publications that uses this dataset.

I. Data collection

I.1. Hardware setup

3 Colibri wireless IMUs (inertial measurement units) were used:
– sampling frequency: 100Hz
– more information on the unit and the sensors inside can be found in: TrivisioColibriwirelessBrochure.pdf
– position of the sensors:
    – 1 IMU over the wrist on the dominant arm
    – 1 IMU on the chest
    – 1 IMU on the dominant side's ankle

HR-monitor: BM-CS5SR from BM innovations GmbH
– sampling frequency: ~9Hz

Companion unit: Viliv S5 UMPC
– Intel Atom Z520 CPU (1.33GHz) and 1GB of RAM
– labelling the different performed activities was done via a GUI running on the Viliv

I.2. Subjects

9 subjects participated in the data collection:
– mainly employees or students at DFKI
– 1 female, 8 males
– aged 27.22 ± 3.31 years
– BMI 25.11 ± 2.62 kgm^{-2}
– for detailed information on each of the test subjects, see: subjectInformation.pdf
– all subjects have agreed to the usage of recorded data for scientific purposes

I.3. Data collection protocol

Each of the subjects had to follow a protocol, containing 12 different activities. The protocol is described in: DataCollectionProtocol.pdf.
Furthermore, a list of optional activities to perform was also suggested to the subjects. The list contained a wide range of everyday, household and sport activities. From this list, in total 6 different activities were performed by some of the subjects in addition to the protocol. A brief description of all the 18 performed activities can be found in: DescriptionOfActivities.pdf.

I.4. Summary of collected data

This is a realistic dataset, therefore some data is missing. Missing data had 2 main reason:
–   Data dropping due to using wireless sensors. This however only occurred very rarely: the 3 IMUs had a "real" sampling frequency of 99.63Hz, 99.89Hz and 99.65Hz for hand, chest and ankle IMU placement, respectively.
–   Problems with the hardware setup, causing e.g. connection loss to the dongles or system crash. Due to these problems, some activities for certain subjects are partly or completely missing.

Altogether, over 10 hours of data were collected, from which nearly 8 hours were labeled as 1 of the 18 activities performed during data collection. A summary of how much data was collected from each of the 18 different activities per subjects is shown in: PerformedActivitiesSummary.pdf.

II. Data format
II.1. Synchronized and labeled raw data from all the sensors (3 IMUs and the HR-monitor) is merged into 1 data file per subject per session (protocol or optional), available as text-files (.dat). Each of the data-files contains 54 columns per row, the columns contain the following data:

–   1   timestamp (s)
–   2   activityID (see II.2. for the mapping to the activities)
–   3   heart rate (bpm)
–   4-20   IMU hand
–   21-37   IMU chest
–   38-54   IMU ankle

The IMU sensory data contains the following columns:
–   1   temperature (°C)
–   2-4   3D-acceleration data (ms^{-2}), scale: ±16g, resolution: 13-bit
–   5-7   3D-acceleration data (ms^{-2}), scale: ±6g, resolution: 13-bit*
–   8-10   3D-gyroscope data (rad/s)
–   11-13   3D-magnetometer data (µT)
–   14-17   orientation (invalid in this data collection)

Missing sensory data due to wireless data dropping: missing values are indicated with NaN.
Since data is given every 0.01s (due to the fact, that the IMUs have a sampling frequency of 100Hz), and the sampling frequency of the HR-monitor was only approximately 9Hz, the missing HR-values are also indicated with NaN in the data-files.

* This accelerometer is not precisely calibrated with the first one. Moreover, due to high impacts caused by certain movements (e.g. during running) with acceleration over 6g, it gets saturated sometimes. Therefore, the use of the data from the first accelerometer (with the scale of ±16g) is recommended.

II.2. Activity IDs:
- – 1   lying
- – 2   sitting
- – 3   standing
- – 4   walking
- – 5   running
- – 6   cycling
- – 7   Nordic walking
- – 9   watching television
- – 10   computer work
- – 11   car driving
- – 12   ascending stairs
- – 13   descending stairs
- – 16   vacuum cleaning
- – 17   ironing
- – 18   folding laundry
- – 19   house cleaning
- – 20   playing soccer
- – 24   rope jumping
- – 0   other (transient activities)

Note: data labelled with activityID=0 should be discarded in any kind of analysis. This data mainly covers transient activities between performing different activities, e.g. going from one location to the next activity's location, or waiting for the preparation of some equipment. Also, different parts of one subject's recording (in the case when the data collection was aborted for some reason) was put together during these transient activities (noticeable by some "jumping" in the HR-data).

References

1. A. Reiss and D. Stricker. Introducing a New Benchmarked Dataset for Activity Monitoring. *The 16th IEEE International Symposium on Wearable Computers (ISWC)*, 2012.
2. A. Reiss and D. Stricker. Creating and Benchmarking a New Dataset for Physical Activity Monitoring. *The 5th Workshop on Affect and Behaviour Related Assistance (ABRA)*, 2012.