7th Framework Programme
Theme 3 "Information and Communication Technologies"
FP7-ICT-2007.8.0 FET-Open

**Project acronym:** OPPORTUNITY
**Project full title:** Activity and Context Recognition with Opportunistic Sensor Configurations
**Grant agreement no.:** 225938
**Project start date:** February 1st 2009
**Project duration:** 36 months

# Deliverable D5.1
# -
# Stage 1 Case Study Report and Stage 2 Specification

**Deliverable Originator:** UNI PASSAU
**Deliverable Contributors:** UNI PASSAU, ETH Zürich, JKU, EPFL
**Deliverable Version:** 0.23
**Deliverable Due Date:** M12 (31.01.2010)
**Deliverable Submission Date:** 31.01.2010, revised 29.03.2010
**Classification:** Public
**Project Co-ordinator:** ETHZ-Dr. Daniel Roggen
**Project Partners:** ETH Zürich - Prof. G. Tröster; University of Passau - Prof. P. Lukowicz; Johannes Kepler Universität Linz - Prof. A. Ferscha; Ecole Polytechnique Fédérale de Lausanne - Prof. J. del R. Millán

# Contents

# Executive Summary

This deliverable describes work that has been done towards the first generation scenarios and demonstration. The original plan was to base the first generation evaluation on simple, easily repeatable experiments consisting of elementary motions and context components (e.g. presence). However, after elaborate discussion, the consortium has decided to change that plan and begin the project with a recording of a large, complex data set that can be used for stage one, stage two, and to a large degree stage three studies.

The data set is based on a breakfast scenario, which is well studied in the literature, rich in terms of different activity types, and can be recorded in a laboratory setting in a realistic way. The recording contains around 25 hours of data from 12 subjects. On the low level there are around 30'000 individual actions (e.g. picking up a knife, opening a drawer). On the highest level (getting up, breakfast preparation) we have around 200 context instances. All of those were annotated during the recording and are currently being verified/re-annotated using the video stream. While the number of high level contexts is not unusual for this type of experiment, the number of annotated low level actions is far beyond what is available in other data sets.

This deliverable starts with an explanation of the above decision and the reasons why existing publicly available data sets are not adequate. We then proceed to elaborate the scenario on which the recording is based, and describe in detail the sensors that were used. We also present a set of tools that was used in the recording (including some tools that were developed in the course of the project). We finish with a short assessment of the quality of the recorded data.

Examples of the actual use of the data can be found in D1, which explains the methods and algorithms developed in the first year of the project.

# 1 Introduction

According to the DoW the first generation demonstrator was supposed to be: "relatively simple examples of basic activity components such as (1) presence and location, (2) modes of locomotion and posture, and (3) hand gestures." Second generation were to "combine several basic components into more complex activities". Finally the third generation was to "built upon complex combination of basic components and demonstrate scenarios clearly motivated by and connected to real life application areas".
For the first generation the original plan was to base the evaluation on simple easily repeatable experiments. However, after elaborate discussion, the consortium has decided to change that plan and begin the project with a recording of a large, complex data set that can be used for stage one, stage two, and to a large degree stage 3. The decision has been based on the following considerations:

- Aiming at the study of dynamic, changing sensor configurations the project requires highly multimodal data sets. Recoding such data sets involves considerable effort, even in small scale experiments. Thus, when making such effort it is desirable to ensure that the resulting data can be used for more then one study.
- As the project evolves it will be desirable to be able to test new techniques against older ones on several different data sets. It would also be good to have data sets that can be made public so that other groups can compare their results against our methods in the future.
- The different stages of scenarios are not independent. Instead, they comprise a hierarchy in which each stage builds on the other. The basic activities which are the topic of the first stage are building blocks of the more complex activities of the second stage which in turn are the components of the complex scenarios in the third stage. Dealing with the hierarchy is an important aspect of the recognition method development. To this end data sets that include a sufficient number of annotated instances of each hierarchy level are needed.

In summary, the strategy of recording a single elaborate data set that can be used through the project will (1) reduce the experimental effort (seen over the entire course of the project), (2) allow more methodologically "cleaner" evaluation and comparison of different methods and (3) and allow us to better include interdependencies between the different activity levels in the methods that we are developing.

## 1.1 Opportunity Data Set Requirements

Recently, publicly available data sets have started emerging in the area of context recognition (see related work below). However, due to the diversity and complexity of the context recognition domain it is difficult to define a few "standard" sets. Instead, there are many aspects that need to be considered in different applications. Given the aims of the OPPORTUNITY project, the following aspects have been identified as essential:

1. Highly multi modal sensor setup. In particular we need numerous sensors that can provide information about each activity (otherwise testing the effect of changing sensor configuration makes little sense). These sensors should be both complementary (providing different type of information about the activity) and redundant with slight differences (providing the same type of information but being nonetheless different enough to require adaptation). To provide

well generalizable results we need broadly different sensor types including different wearable and environment integrated devices.

2. Multilevel activity set. As described above the data set is intended to combine work on different levels of activity: from simple basic gestures to complex compound activities (see fig 1). This hierarchy has to be adequately reflected in the data set. The main problem here is annotation. While, trivially, any data set containing compound activities also contains the basic building blocks, in most datasets focused on high level activity the low level components are not annotated (see related work below).
3. Large number of instances. Solid statistical evaluation requires a large number of training and testing instances of each activity. As described in D1, some approaches envisioned by the project require a second round of training (to be able to utilize new sensors) which further increases the amount of instances that are needed.
4. Significant number of different users which is needed to allow us to study how our methods perform in user independent scenarios.

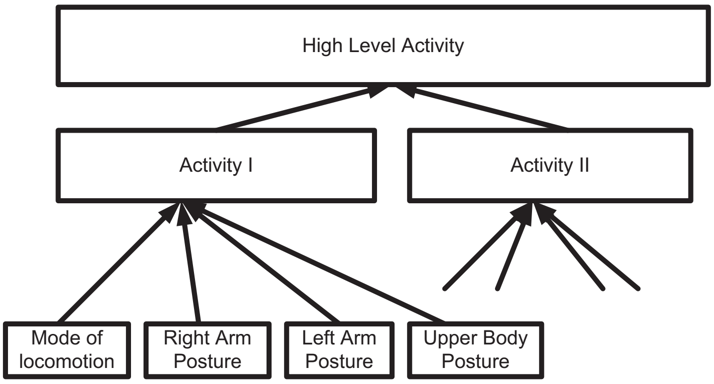



Figure 1: Schematic Example of an activity hierarchy. Mode of locomotion, postures of left and right arm and upper body posture distinguish activity I. Activities I and II together form the high level activity.

## 1.2 Existing Datasets

In literature a number of public activity recognition data sets are presented. However, as shown below, there is no data that combines **all** of the OPPORTUNITY requirements.

### 1.2.1 PlaceLab

The most popular data set available in pervasive / ubiquitous area is the so called PlaceLab data set (see [3]). Longtime data recordings with a rich multimodal sensor environment captures the behavior and activities of test subjects over days or weeks in a sensor equipped apartment. Environment sensors (like temperature, or humidity sensors) capture the environmental conditions of the living area. Sensors attached to objects collect information about object interactions. In the beginning only 3 acceleration sensors capture on body posture and mode of locomotion, most information has been added in offline annotation sessions looking at the video stream or listening to audio recording. Only one data stream from each set of cameras and set of microphones have been recorded according to the current position of the person. This data set provides a rich set of object interactions for behavior research and data for context algorithms. Neither specific and well defined gestures nor a high number of repetitions of gestures is the goal of this project. Capturing a single gesture with several sensor modalities also had a lower priority.

### 1.2.2 Kitchen data set

Data recording in a kitchen environment has been performed by a group from TU Munich (see [7]). They focus on marker free motion capture of complex gestures. The data set provides video, motion capture, RFID reader and reed switch information. RFID reader and reed switches give timing information when the subject interacts with the kitchen environment. There have not been any on body sensors like acceleration or gyroscope sensors capturing body postures or modes of locomotion.

### 1.2.3 Activity Recognition in a home setting

Another data set has been presented in [9]. The authors recorded for over a month the test subject's life. Digital or binary sensors (*idle* or *active*) like reed switches give information when the person interacts with furniture or objects of interest. Neither video, audio, modes of locomotion nor posture information have been recorded. The dataset's weak point is the reduced number of sensor systems and of test subjects.

## 1.3 Opportunity Data Set Summary

The set contains around 25 hours of data from 12 subjects. On the low level there are around 30'000 individual actions (e.g. picking up a knife, opening a drawer). On the highest level (getting up, breakfast preparation) we have around 200 context instances. All of those were annotated during the recording and are currently being verified/re-annotated using the video stream. While the number of high level contexts is not unusual for this type of experiment, the number of annotated low level actions is far beyond what is available in other data sets. On the other hand, the availability of annotations for all low level activities is crucial for the development of complex, hierarchical recognition methods.
The experiment was carefully designed to provide realistic data. To this end the subjects were given loose high level instructions with respect to the activities and a good approximation of a real life environment was established. Nonetheless, this is clearly an artificial data set recorded in a laboratory setting. On the other hand, by choosing such a setting we were able to get a large number of repetitions of the same activity with the ability to annotate each individual instance. Both is difficult when recording in real life where people are free to do whatever they like and neither permanent observer presence nor detailed video recording are possible.

# 2 Case Study breakfast scenario

The data set requirements include a high number of instances of different gestures recorded by a high number of on-body, environmental and object-attached sensors. A breakfast-related scenario has been chosen as it has extensively been used in literature (for example in [11],[10],[4] or [8]). The tasks of the scenario are everyday life activities and well known. This reduces the training time. The test environment is usually available in every house.

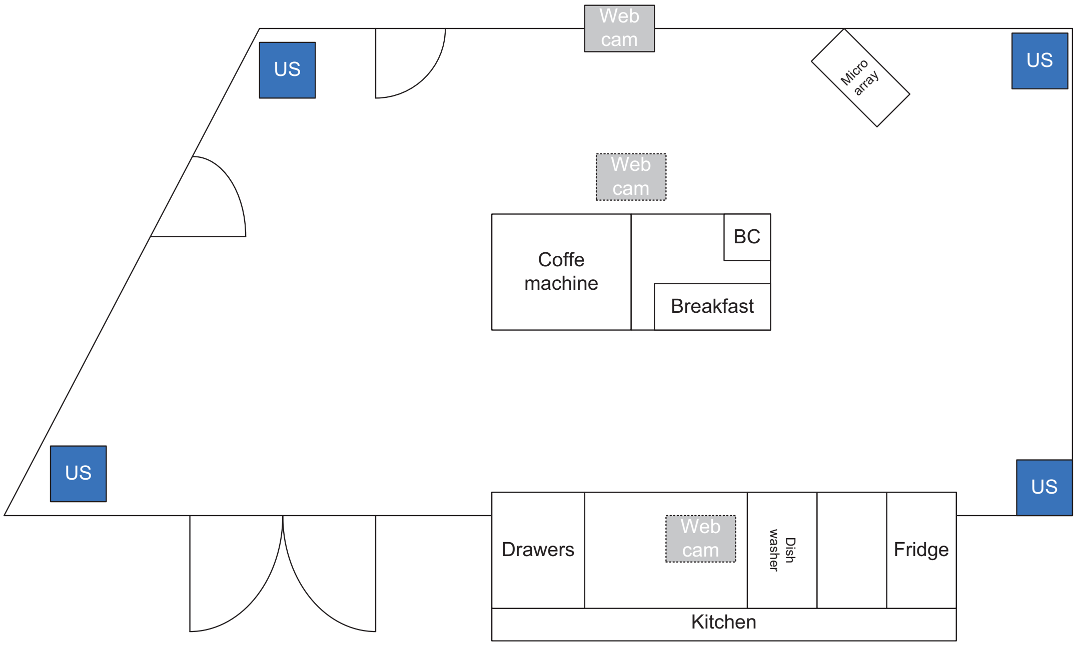



Figure 2: Room layout of the test environment. The Ubisense receiver parts (blue squares, US) have been installed in the corners of the room, the microphone array on the upper side of the room. The dashed webcams have been attached to the ceiling. The bread cutter (BC) is on the table in the center of the room.

## 2.1 Description of the scenario

The experiment has been set up in a room (see figure 2) sized 8x5x3m. The room has 3 doors, a kitchen section and a table which is placed in the center of the room. We divided the case study into two parts both providing a high number of atomic instances:

## 2.2 Drill Session

A highly scripted session has been introduced to satisfy the dataset requirements. The test subject sequentially has to go through a sequence of gestures (20 repetitions):

1. Open and close the fridge (2 activities)
2. Open and close the dishwasher (2 activities)
3. Open and close 3 drawers (each at different heights) (2 activities each)
4. Open and close door 1 (2 activities)
5. Turn on and off the lights (2 activities)
6. Open and close door 2 (2 activities)
7. Clean table (1 activity)
8. Drinking (standing) (2 activities)
9. Drinking (sitting) (2 activities)

The activities have been chosen to be recognized by a combination of environmental sensors, body worn sensors and sensors attached to objects. For each run we therefore record 21 different activities.

## 2.3 Real Activities

The second part of the case study provides gestures in a more natural way. The test subject goes through certain well defined and well known higher level activities. On the one hand this part provides additional (but more natural as non scripted) instances of atomic activities on the other hand it provides higher level activities as well (as for example *laying plates*, *preparing sandwiches*, *preparing coffee* or *drinking water*).

Remember that an atomic activity - for example taking a plate out of the drawer - usually consists of several low level gestures like reach plate, take plate, move plate.

### 2.3.1 Getting up

At the beginning the person lies on a canvas chair and relaxes. After a while the person gets up, opens a door and goes out for a walk. The test person walks 100 m flat then downstairs - some meters flat again turns around, walks upstairs and then back to the door. After reentering the room she enters the room and closes the door.
Atomic gestures in this part are:

- Getting up
- Open the door 1
- Close the door 1
- Open the door 2
- Close the door 2
- Walking
- Walking downstairs
- Walking upstairs

### 2.3.2 Breakfast preparation

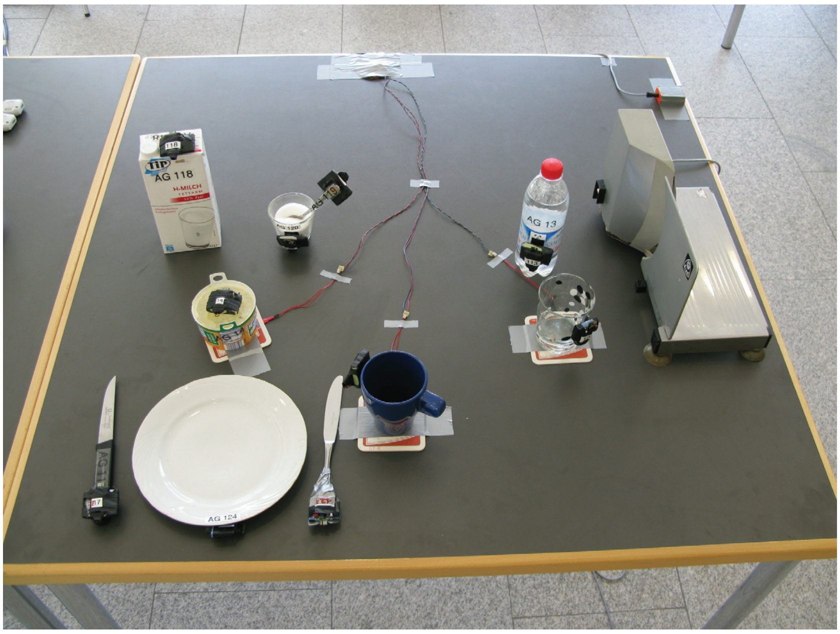



Figure 3: Part of the table with the breakfast tableware and the bread slicer.

The subject starts preparing the breakfast. Therefore she gets the plate, knife and spoon from different drawers in the kitchen. The bread is stored in a drawer, the salami, milk, water bottle and

cheese are located in the fridge. The glass, cup and sugar are located in a rack above the kitchen. She then prepares a cup of coffee with the coffee machine, adds coffee and milk and puts the cup on specific position(definded by the specific force sensor attached to the table).
The sandwich preparation involves cutting two slices of bread and salami with the bread cutter. The spread cheese is put on the slices with the knife and on top of this the subject puts the salami. The sandwiches are placed on the plate.
In the end the test subject pours water into the glass. She can always open or close one of the two possible doors of the room. Figure 3 depicts a typical breakfast setup.
A subset of typical atomic activities of this section are for example:

- Walking
- Open the door 1
- Close the door 1
- Open the door 2
- Close the door 2
- Open drawer 1
- Close drawer 1
- Open drawer 2
- Close drawer 2
- Open drawer 3
- Close drawer 3
- Open fridge
- Close fridge
- Reach item

### 2.3.3 Breakfast

After preparation the subject sitts down and starts eating and drinking. From time to time she gets up and open or closes doors.
Typical atomic activities of this section are for example:

- Open the door 1
- Close the door 1
- Open the door 2
- Close the door 2
- Reach item
- Move item
- Sitting
- Drink
- Eat
- Pouring Water
- Coffee cup
- Water glass

### 2.3.4 Cleaning up

After finishing the breakfast the subject cleans up. She puts the food back to the original locations, places the dishes in the dish washer and wipes the table with a towel. From time to time she opens or closes doors.
Typical atomic activities of this section are :

- Walking
- Open drawer 1
- Close drawer 1
- Open drawer 2
- Close drawer 2
- Open drawer 3
- Close drawer 3
- Open fridge
- Close fridge
- Open dish washer
- Close dish washer
- Reach item
- Move item
- Release item
- Wipeing

### 2.3.5 Going back to sleep

After these activites the subject closes the doors turns off the lights and lies back on the chair.
Typical atomic activities of this section are for example:

- Walking
- Open the door 1
- Close the door 1
- Open the door 2
- Close the door 2
- Turn on light
- Turn off light
- Standing
- Lying down

## 2.4 Richness of the Scenario

As described in the section 2.3 the scenario provides a huge number of different gestures and for each of these gesture a high number of repetitions. In addition to this richness the records have also overlapping activities like walking and moving of items at the same time or moving items and closing of doors. Especially when working with acceleration sensors these overlapping and simultanious occuring activities put additional complexity to activity recognition algorithms.

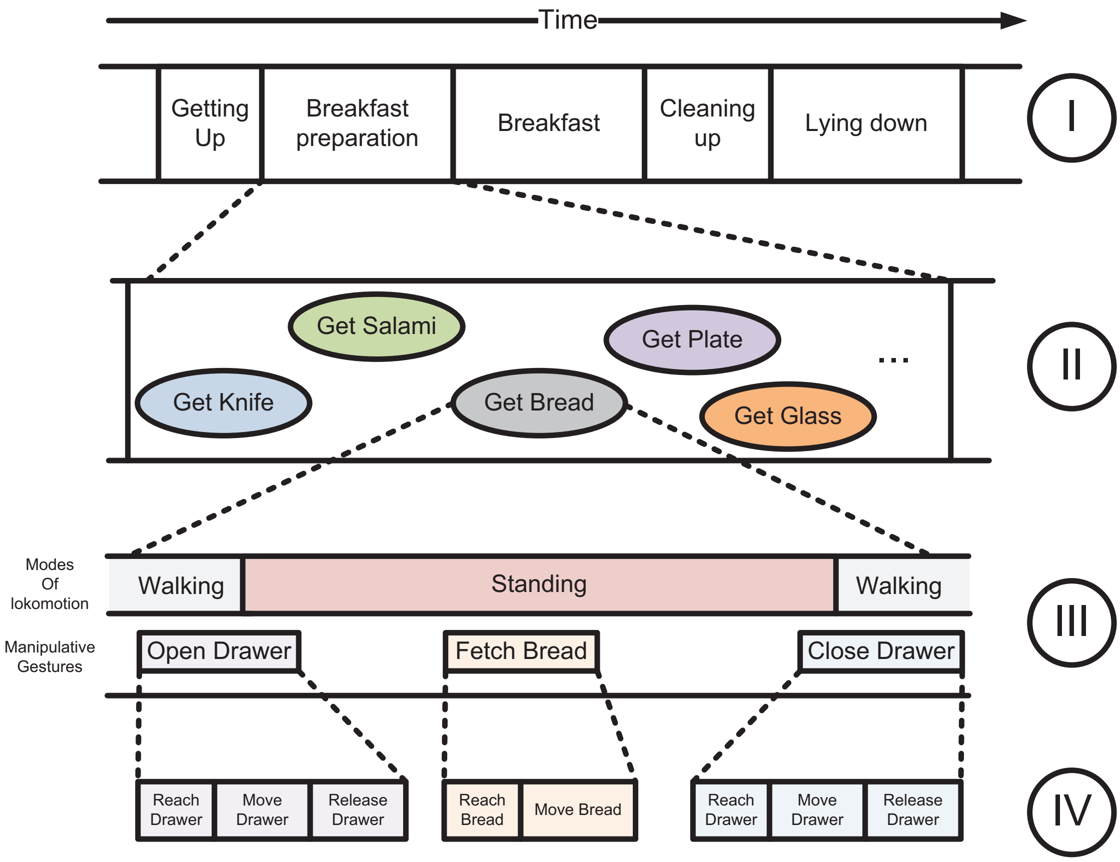



Figure 4: **Temporal decomposition of activites**. Level I is the highest activity level available in the setup. Level II zooms in into one high level activity, in this level the activities are not temporal ordered and depend on the execution sequence of the testsubject. Logical, physiological and spatial limitations distinguish the order of activites in Level III. Here the activities are modes of locomotion and manipulative gestures. Level IV encapsulates the atomic gestures forming the manipulative gestures of level III.

Figure 4 depicts the temporal decomposition of activities at different temporal zoom levels. Level I can be compared to the description presented in section 2.3 or the scripted sequence of the drill session. The temporal sequence of these activities is static. If we pick out one of these high level activities and look at it more closely this activity is distinguished by lower level (but still complex) activities (symbolized as ellipses) on level II. The order of these activities is not fix and differs according to the execution order of the different test subjects. Zooming in on level III shows that the activities of level II are dominated by modes of locomotion (for example walking, standing, sitting) and by manipulative gestures (like moving, reaching, grasping or releasing). We want to point out that it is possible that manipulative gestures and modes of locomotion overlap. Logical, physical and spatial limitations distinguish and influence the orders of these activities.

### 2.5 Experimental Protocol

Before we started the experiments we prepared the room and equipped the kitchen with the used sensors. The Ubisense ultra wide band localization system has been evaluated at first to see whether there are interferences in the environment degrading the localization accuracy. Thus we measured therefore with 6 tags 15 positions each (exact coordinates measured with a laser meter with sub cm accuracy) at different heights. The accuracy of the localization system is in the range given by the company (sphere with a diameter of 20 cm to 30 cm).
There are 7 computer involved in collecting the data. Before starting each computer is NTP time synchronized with a local server. Table 1 holds information about the data collection tasks of the used computers. The sensors are attached to the body and objects, we keep the correct orientation

| Computer | Connected Sensors |
|---|---|
| Laptop | Bluetooth Accel. Sensors, Magnetic coupling, Xbus (on body) |
| PC | Xsens Table and Chair, Powersensors, Ubisense |
| Laptop | Audio (Mic. Array and onbody wireless micr.) |
| Laptop | Inertial Cube and Sunspots |
| Laptop | Video Data streams |
| Laptop | USB Accel Sensors(Environment) |
| Laptop | Reed Switches(Environment), accel and gyro sensor (Object) |

Table 1: 7 Computers capture the data streams of the used sensor systems.

of the sensors in mind. Figure 5 gives an overview of the on body locations of the sensor systems attached to the subject.

The first run of the experiment can be compared to a training session where the instructor explains the tasks and the sequence of the activities. A usual run takes 15 to 25 minutes. After each session (5 runs and 1 drill session) we carefully replaced the batteries of the sensors. In order to have a synchronization possibility the subject has to perform a special gesture. The person claps and we can use this gesture to synchronize video and audio datas treams with the on body sensors as the clapping is easy to spot either in video and audio data streams and also in on body sensor data streams (Example see figure 6).

During the runs there are several persons involved in labeling the current activities. Each person labels the synchronization gesture. One person is reponsible for labeling modes of locomtion (*standing*,*walking*,*sitting*), three other persons have to label different level of activities (high level activities like *Preparing breakfast*, midlevel activities like *Slicing Bread* or lowlevel activites like *Moving Bread*).

## 3 Activity Sensor Mapping

In this section we present how gestures can be captured with different sensor modalities. We distinguish between sensors attached to any body parts, sensors embedded in the environment and sensors attached to objects.
For a detailed information about the used sensor systems see section 4. We use so called MARG

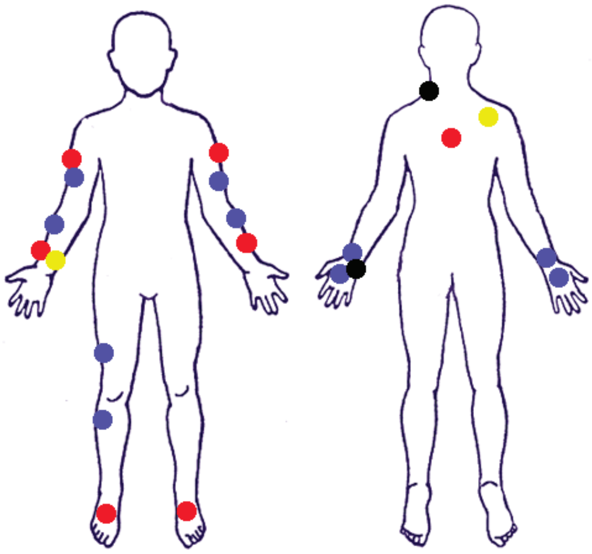


Figure 5: The sensors are attached to the body as shown in the schematics. The circles correspond to different on body sensor systems. The red circles are MARG sensors, the yellow circles correspond to the magnetic sensor system. Bluetooth connected acceleration sensors are depicted as blue dots. The wireless microphones are black circles. Not depicted are the four shoulder attached Ubisense localization tags.

units ( Magnetic Earth Field, Rotation and Gravity + Acceleration, XSens, Sun Spot and Intertial cubes), audio, an oscilating magnetic field sensor ([5]) and acceleration / gyroscope sensors for on body motion capturing. Environmental sensors embedded in doors, attached to drawers, fridge, dish washer, chair, table or electric devices (bread cutter and coffee machine) are acceleration / gyroscope based sensors, MARG units, reed switches, video, audio or an ultra wide band based localization system (Ubisense) or force resistive sensors and power sensors. Sensors attached to objects have been acceleration / gyroscope based sensors. We want to point out that there are several sensor systems with only acceleration / gyroscope combinations differing from the MARG definition.

This high number of different sensor modalities gathers information about gestures from different sensor modalities and from different perspectives. In the following subsections we present different situations and how the used set of sensors give information to recognize gestures in this situation.

### 3.1 Sipping from the coffee cup

Sipping from the coffee cup has certain distinct properties: The person usually stands or sits (modes of locomotion), holds the cup in the hand and moves the hand near the mouth. After drinking, the cup is put back to the table. There are certain sensor combinations which give information about body / arm posture, modes of locomotion and object interaction:

**On body sensors** : MARG units attached to the subject's arm at different positions give information about the posture of the arm. Acceleration and gyroscope sensors measure the occuring acceleration and rotation values plus the gravity ratios at different sensor axis. Relative position (distance and angle representatives) information between chest and hand wrist in addition to wrist orientation is derived from the oscillating magnetic field system describing

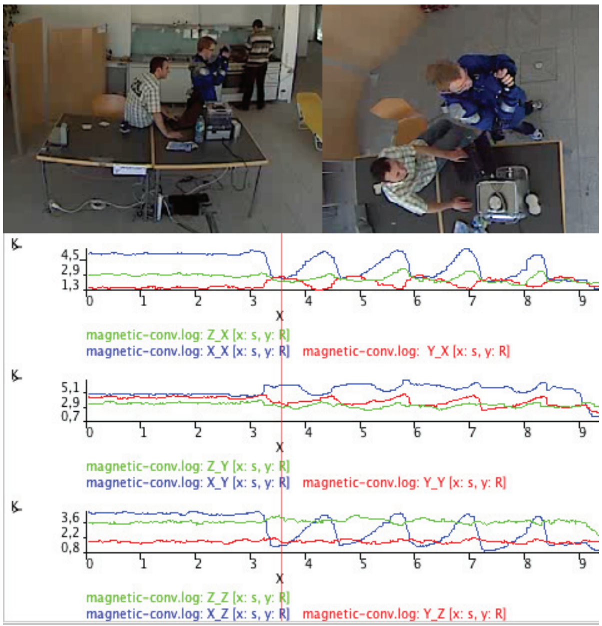



Figure 6: A screenshot from the labeling tool during a synchronization gesture. The data stream is taken from the oscillating magnetic field sensor. The images are taken from the side and ceiling webcams.

the arm posture in a different way compared to MARG or acceleration / gyroscope based sensors. MARG Sensors on the shoes capture the current mode of locomotion, together with upper body acceleration sensors and acceleration sensors attached to the knee it is possible to determine whether she is walking, standing or sitting.

**Environmental sensors** : Video and audio based localization determine the position of the person. The ultra wide band based localization system can also be used to distinguish between standing walking and sitting: We attached the tags at the shoulders of the person, if she sits down, the Z component (height) of the subject's coordinate changes noticeablly. A MARG unit attached to the chair detects interaction with the chair, vibrations on the table give distinct signals on acceleration part of another MARG Unit. Force resistive sensors measure when the person takes the cup and when she puts it back on the table giving a rough time interval when to spot gestures on on body sensor signals.

**Object embedded sensors** : An acceleration and gyroscope combination attached to the coffee cup captures movements and orientation changes while the person is interacting with the cup.

## 3.2 Bread Slicing

When the person is using the bread slicer to slice the bread there are certain sensors available capturing this event. Proximity information (*The subject is next to the bread slicer*) and current mode of locomotion (Usually persons slice bread while standing) limit the possibilities of activities. Power sensors attached to used devices give a good estimation of the time frame.

**On body sensors** : MARG units and acceleration, gyroscope provide information about the current mode of locomotion and the posture information about how the upper body part is oriented. Arm attached MARG sensors and the oscillating magnetic field sensor capture arm position and the arm movement. Microphones at the hand and near the neck provide environmental sound information and allow to distinguish different sounds by their intensity and their frequencies.

**Environmental sensors** : The microphone array allows to estimate the position of the sound source and can also be used to classify the sounds. Video based localization and ultra wide band based localization limit the current position of the person to a certain area. The power sensor attached to the bread slicer captures the current power consumption. It has been shown that the power consumption gives a good indication of the current state of the connected device(see [2]). MARG units on the table sense vibration generated by the bread cutter.

**Object Embedded Sensors** : Object interactions (taking the bread, cutting the bread, releasing the bread) are sensed by acceleration and gyroscope sensors on the bread.

### 3.3 Taking Milk out of the fridge

An example where an environmental and object attached sensors contribute to the classification process is the high level event of taking a bottle of milk out of the fridge.

**On body sensors** : MARG units and acceleration and gyroscopes attached to the arm and to the upper part of the body capture body posture, the oscillation magnetic field sensor gives different (but also useful)posture information compared to MARG or acceleration gyroscope combinations. The microphone attached to the forehand can detect when the door is opened as this opening sound is very specific. Standing (mode of locomotion) is detected by Acceleration / gyroscope sensors and by the MARG units.

**Environmental sensors** : Video and ultra wide band localization captures proximity information (person is next to the fridge). Reed switches attached to the fridge door captures opening and closing events giving a rough time frame. Acceleration and gyroscope information of attached sensors at the fridge door help to recognize this opening and closing events.

**Object Embedded Sensors** : Opening and closing of the fridge door is captured by acceleration and gyroscope sensors attached to the water bottle and milk box.

Notice that the presented examples show that performed gestures are captured by different sensor modalities and there are always at least two systems contributing to the classification process.

## 4 Sensors

In this section we describe the rich set of sensors being used in the setup. As explained earlier we used sensors on body, attached to objects and integrated in the environment to capture interactions, activities and gestures.

## 4.1 On-body Sensors

Estimating modes of locomotion and postures usually involves attaching sensors to the body. We present the used sensor modalities in this subseciton.

### 4.1.1 Motion Jacket

**Description** The Motion Jacket is a garment (see Figure 4.1.1) containing five Xsens inertial units: one acts as a reference and is mounted at the middle back of the jacket, while the other four are mounted respectively at the upper and lower parts of the two arms. Each unit contains an acceleration sensor, a gyroscope and a magnetic field sensor (each of these along three axis). The units communicate with a central coordinator device through a serial bus and the device is connected to a PC via USB or Bluetooth. The coordinator is powered by four standard AA rechargeable batteries.

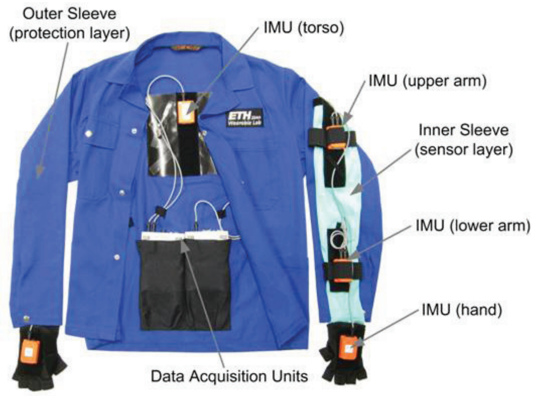



Figure 7: Motion Jacket: 5 inertial measurement units embedded in garment.

**Recording conditions**

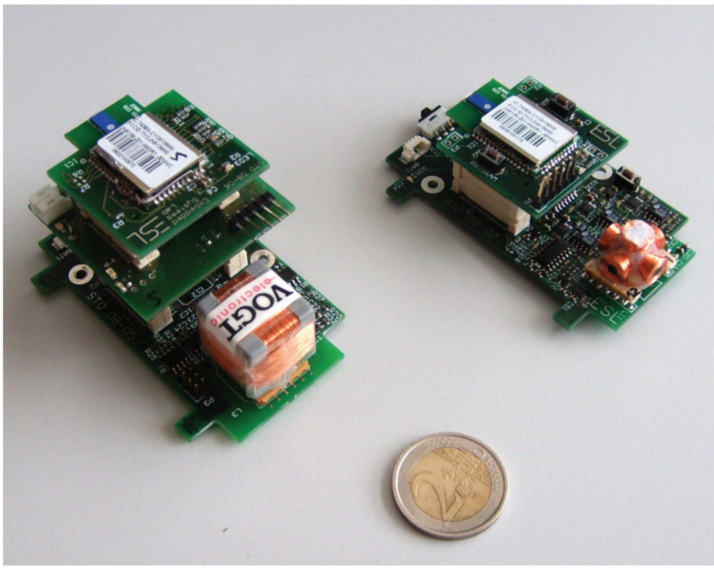
- On-body location: mid-back, lower and upper arms
- Sampling frequency: 30Hz
- Connection: USB
- Client side: CRN Toolbox

### 4.1.2 Magnetic Coupling Based Sensor

**Description** The magnetic coupling based sensor system provides relative position information. We attached the fieldtransmitter at the back of the testsubject, the receiver part of the system was attached to the main arm of the person.

**Recording conditions**

- Sampling frequency: 87 Hz
- Connection: USB (RS232 to USB)
- Client side: CRN Toolbox
- Max distance: 85 cm
- On body Location: transmitter on back of person (scapula), receiver at wrist of main arm



**Signal Post Processing** The logged raw values (see appendix for details) cannot be used for further processing without conversion.
The transmitter part generates an oscilating magnetic field at 3 perpendicular transmitter axis in sequence. The receiver part attached to the wrist holds 3 three axis receiver coil. For each transmitter axis 3 adc values are estimated (measuring the induced voltage on each receiver axis.). As the magnetic field attenuates very quickly, the signal has a high range (several volts in the near, and some mV at 80 cm). We solve this high dynamic range by the use of adjustable amplifiers. The processed values $p_i$ of the magnetic field sensor are calculated by using the ADC value $ADC_i$ and the corresponding amplification value $AMP_i$:

$$p_i = ADC_i \cdot 2^{7-AMP_i} \tag{1}$$

The processing has to be done with all 9 ADC values (and corresponding AMP values).
In order to eliminate outliers, a Median filter with a sliding window size of 10 samples is used for each of the 9 channels. A measurement $\mathcal{M}_i = \{x_x, y_x, z_x, x_y, y_y, z_y, z_x, z_y, z_z\}$ hold information about the induced voltages at the receiver axes while a transmitter axis generated the field. $i_j$ is the induced voltage at receiver axis $i$ while transmitter axis $j$ generated the field.

**Description of the Signal** Each line of the log file gives a relative position estimation (together with orientation of the arm) between the test subject and the wrist.

- $mag = \sqrt{x_x^2 + y_x^2 + z_x^2 + ... + z_z^2}$ gives a good approximation of the distance between the arm and the back of a person. If the signal is high, the hand is very near to the back (if it is low - the hand is far away)
- $s_x = \frac{\sqrt{x_x^2+y_x^2+z_x^2}}{\sqrt{x_y^2+y_y^2+z_y^2}}, s_y = \frac{\sqrt{x_y^2+y_y^2+z_y^2}}{\sqrt{x_z^2+y_z^2+z_z^2}}, s_z = \frac{\sqrt{x_x^2+y_x^2+z_x^2}}{\sqrt{x_z^2+y_z^2+z_z^2}}$ is a angle representative
- $r_x = \frac{\sqrt{x_x^2+x_y^2+x_z^2}}{\sqrt{y_x^2+y_y^2+y_z^2}}, r_y = \frac{\sqrt{y_x^2+y_y^2+y_z^2}}{\sqrt{z_x^2+z_y^2+z_z^2}}, r_x = \frac{\sqrt{x_x^2+x_y^2+x_z^2}}{\sqrt{z_x^2+z_y^2+z_z^2}}$ are representatives for the orientation of the hand with respect to the transmitter.

The position of the hand in the magnetic field is distinguished by $(mag, s_x, s_y, s_z)$.

### 4.1.3 Bluetooth Acceleration Sensors

**Description** A set of 12 acceleration sensors were placed on the following body parts:

- Left upper arm up
- Left upper arm down
- Left wrist
- Left hand
- Right upper arm up
- Right upper arm down
- Right wrist
- Right hand
- Right knee up
- Right knee down
- Right hip
- Back

The goal in having pairs of sensors in similar positions is to be able to simulate a range of intermediate sensor placements. The algorithms that will be developed within this project will have to show robustness with respect to what would happen in a realistic situation, where it cannot be guaranteed that users wear a certain sensor in a very well defined position. Each sensor board hosts a triaxial accelerometer and has a Bluetooth interface that allows the signals to be recorded wireless through the CRN Toolbox. The sensors were already used in previous research (see for example [1]).


Figure 8: Acceleration sensors mounted on .

**Recording conditions**

- Location: 12 body parts
- Sampling frequency: 64Hz, 32Hz (see following text)
- Connection: Bluetooth
- Client side: CRN Toolbox

For the last five sessions, some measures were taken to try to reduce packet traffic on the Bluetooth channels and to improve the data acquisition. To the first end, sampling frequency was thus decreased from 64Hz to 32Hz and only uncalibrated data were sent. Furthermore, the packet counter size was increased from 8 to 16 bits. This allows to detect interruptions up to 65535 samples (34 minutes at 32Hz) without ambiguities, whereas the 8 bit counter allowed only to detect 255 samples (8 seconds).

### 4.1.4 Sunspots

**Description** Sun SPOTs (Sun Small Programmable Object Technology) [?] are wireless sensor network motes developed by Sun Microsystems. A Sun SPOT consists of (i) a processor board (180 MHz, 32 Bit, 512K RAM, 4M Flash) with an on-board 2.4 GHz IEEE 802.15.4 radio and USB

inteface, (ii) a sensor board with different sensors (2G/6G three-axis accelerometer, temperature sensor snd light sensor), interfaces (six analog inputs, five general I/O pins, four high-current output pins) and eight tri-color LEDs on top of the sensor board, as well as (iii) a 3.7V rechargeable 750mAh lithium-ion battery. Sun SPOTs are powered by a small-footprint Java virtual machine called Squawk, which can host multiple applications concurrently and does no require an underlying operating system. The Sun SPOT SDK is currently available in version 5.0, and it also includes code samples and an emulator. The entire Sun SPOT project, hardware, operating environment, Java virtual machine, device drivers and applications, are available as open source.
We used the three-axial LIS3L02AQ accelerometer [**?**] mounted on the sensor board of the Sun SPOT. The z-axis is perpendicular to the Sun SPOT boards, the x-axis of the accelerometer is parallel to the row of LEDs on the sensor board and the y-axis is parallel to the long edge of the sensor board. The acceleration for each of the three axes ($a_x$, $a_y$ and $a_z$), the total acceleration $|\bar{a}| = \sqrt{a_x^2 + a_y^2 + a_z^2}$ as well as the inclination (tilt) $\Theta_{axis} = arcsin(a_{axis}/|\bar{a}|)$ in radians of each axis with respect to the total acceleration the SPOT is experiencing have been recorded.


Figure 9: Sun SPOT acceleration sensor mounted on a shoe.

**Recording conditions**

- On-body Location: feet (below the right/left ankle on the right/left shoe)
- Sampling frequency: $\approx$ 10-35 Hz (see text below)
- Connection: IEEE 802.15.4 radio (2.4 GHz)
- Client side: Proprietary Java client based on the Sun SPOT SDK v4.0

One Sun SPOT sensor has been taped to each shoe right below the outer ankle of the foot (see Figure 9). The SPOTs were powered by the integrated 3.7V battery, and the signal of each SPOT was transmitted wirelessly to a dedicated USB receiver which was plugged to a laptop computer. Therefore, the application consisted of two parts, one running on the SPOT and another one running on the laptop computer. For the former, just the reset button of the SPOT had to be pressed in order to start it. As with the Inertiacube3 sensors described in Section 4.1.5, the laptop was carried along to guarantee a constant signal quality whenever the person went outside of the room. We changed the SPOT after each session to recharge its battery, which however sometimes run empty before the end of the session. For several reasons, mainly because of wireless communication problems and manual adjustments to cope with battery life time issues, the sample frequency was not constant but varied between 20 and 30 Hz throughout different sessions.

### 4.1.5 Inertiacube3

**Description** Taking feet orientation and acceleration of the human walk cycle as a reference for determining a users gait two Intersense Wireless Inertia Cube 3 sensors were mounted at the toe box the of shoes (see Figure 10). These sensor units include gyroscope, magnetometers and accelerometer with respect to gravity for 3DoF acceleration, angular velocity and orientation updates at a maximum 180Hz rate.

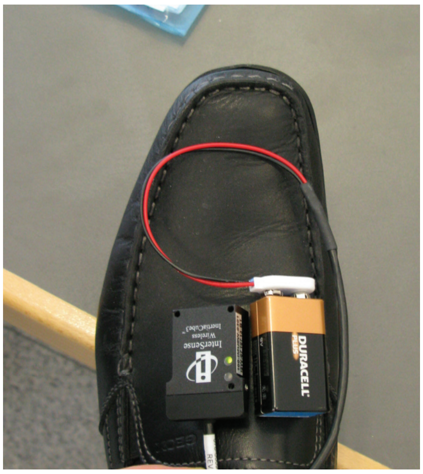




Figure 10: Inertiacube3

**Recording conditions**

- On-body Location: on left and right shoe toe-box
- Sampling frequency: 40 Hz
- Connection: propietary Inertia Cube radio protocol(2,4 Ghz band)
- Client side: Proprietary Java client using Interia Cube SDK dynamic library (dll) with JNI wrapper in Windows XP.

**Remarks:** The sensors are powered by 9V block batteries and were taped together to the shoe upper toe box. The signal was transmitted to the laptop using a USB-powered receiver. Whenever the person went outside of the room to go for a walk, the laptop was carried along to guarantee a constant signal quality. At some early recordings the sensor got disconnected due to the power connector that was not compatible with Duracell battery pins and caused. A new (non-rechargeable) battery lasts for one complete session about 4-5 hours.

## 4.2 Object Sensors

### 4.2.1 Bluetooth Acceleration+Gyroscope Sensors

**Description** A set of 12 objects (cup, glass, 2 knives, plate, bread, salami, milk, water, sugar, spoon, cheese) have been instrumented with a sensor each (see Figure 4.3.1). Each sensor board hosts a triaxial accelerometer and a biaxial gyroscope. The sensors have a Bluetooth interface that allows the signals to be recorded wireless through the CRN Toolbox and have a Li-ion battery that allows for 6-8 hours of recording. The sensors were already used in previous research (see for example [1]).


Figure 11: Acceleration and Gyroscope sensors mounted on sugar, spoon and milk.

**Recording conditions**

- Location: 12 objects
- Sampling frequency: 64Hz, 32Hz (see following text)
- Connection: Bluetooth
- Client side: CRN Toolbox

For the last five sessions, some measures were taken to try to reduce packet traffic on the Bluetooth channels and to improve the data acquisition. To the first end, sampling frequency was thus decreased from 64Hz to 32Hz and only uncalibrated data were sent. Furthermore, the packet counter size was increased from 8 to 16 bits. This allows to detect interruptions up to 65535 samples (34 minutes at 32Hz) without ambiguities, whereas the 8 bit counter allowed only to detect 255 samples (8 seconds).

## 4.3 Environment Sensors

### 4.3.1 USB Acceleration Sensors

**Description** Eight USB acceleration sensors were mounted on parts of the kitchen in order to capture interactions of the test subject with the environment. These sensors feature a triaxial accelerometer and were used in previous work (see for example [6]).

**Recording conditions**

- Ambient location: fridge and dishwasher doors, drawers, kitchen doors, lazy chair
- Sampling frequency: 98Hz
- Connection: USB
- Client side: GenericReader (ad-hoc software)

### 4.3.2 Ubisense

**Description** Ubisense is a wireless ultra wide band system to detect location of tags. The system runs on a dedicated network, where UDP packets are sent, containing the tag positions. In order to estimate the test subject's position we attached 4 tags at the subject's shoulders. The four receivers were placed at the corners of the room (see Figure 4.3.2). We performed tests to assess the precision with which the tag location could be estimated in our setup, resulting in 20cm accuracy.

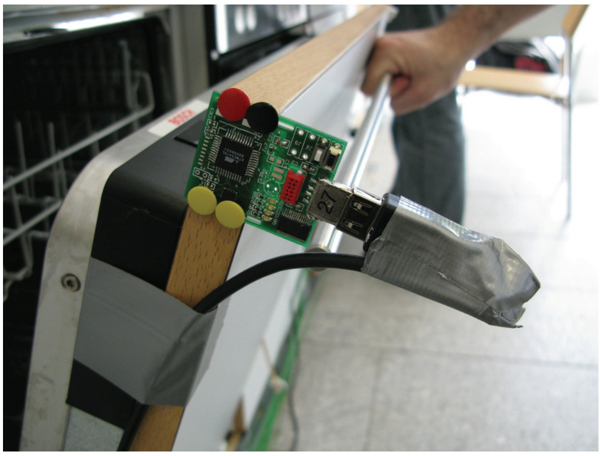


Figure 12: USB Acceleration sensor mounted on dishwasher door.

Figure 13: Ubisense receivers at the room's corners. An accuracy of 20cm was achieved in tag position estimation.

- Sampling frequency: strongly depends of the performance of the Network, changes from subject to subject
- Connection: Network
- Client side: CRN Toolbox, .Net Ubisense to Netwerk Programm

### 4.3.3 Reed Switches

**Description** Magnet-reed switch pairs (see Figure 4.3.3) were used to detect opening and closing of the fridge door, the dishwasher and three drawers. Each time a magnet is in close proximity with a switch, the switch is closed. The signals were recorded by a USB Data Acquisition Module (DAM) which hosts 16 digital inputs. A digital "1" is recorded from the DAM whenever a switch is closed. On most of the objects (apart from the upper drawer) three magnet-switch pairs were installed, in order to detect various levels of opening (barely open, partly open, completely open). On the upper drawer one switch was mounted with three magnets, thus a complete opening of that drawer results in three pulses on the same data line. See Figure 4.3.3 for a detailed placement description.

**Recording conditions**

- Ambient location: fridge and dishwasher

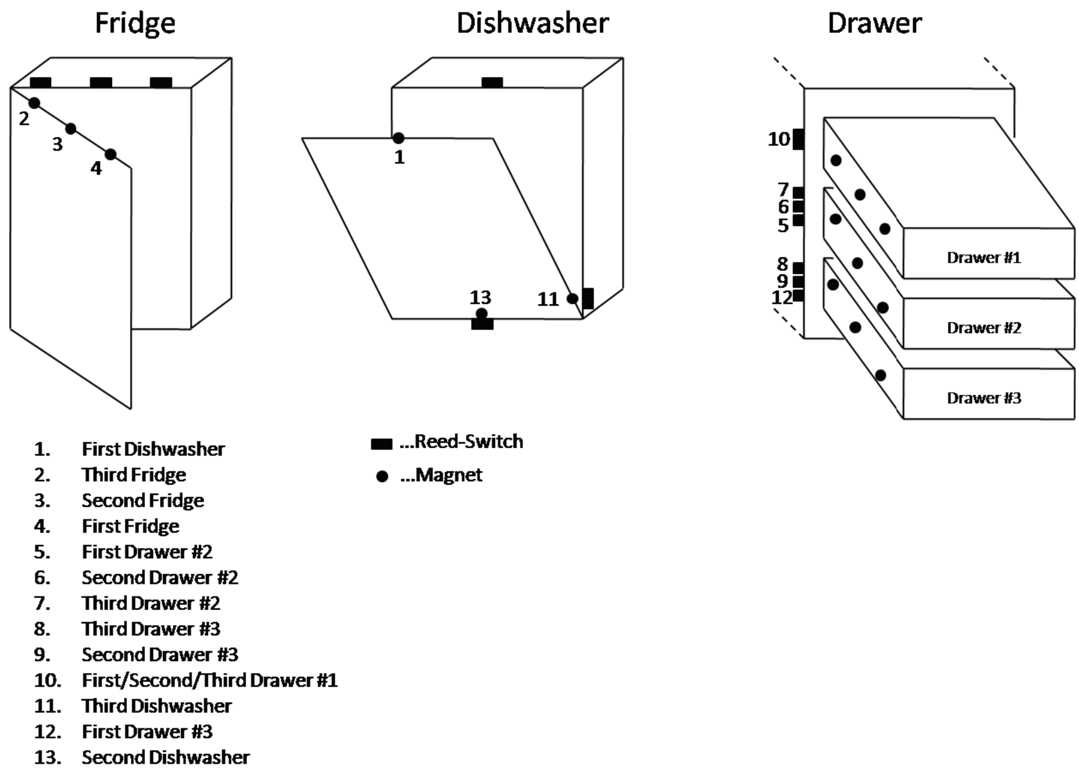



Figure 14: Magnet/Reed Switch pairs placement on various environment items.

doors, drawers (see Figure 4.3.3)
- Sampling frequency: 100Hz
- Connection: USB
- Client side: CRN Toolbox

### 4.3.4 Force Resistive Sensors

**Description** The Force Sensing Resistor sensor changes it's resistance according to the force applied to it. This allows to measure the weight of the item which is on the sensor and whether it is on the sensor or not. 4 Channels can be connected to the sensing module. Figure 2 shows the 3 sensors below objects.

#### Recording Conditions

- Sampling frequency: 48 Hz
- Connection: USB/Zigbee
- Client side: CRN Toolbox

### 4.3.5 Xsens

**Description** Two XSens inertial units (MARG)(see also Section 4.1.1) were mounted respectively underneath the chair and on the table (see Figures). These sensors can contribute in the recognition of activities involving for example food preparation.

#### Recording conditions

- Location: chair, table

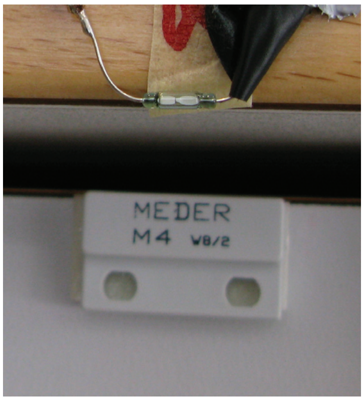



Figure 15: Example of a Magnet-Reed Switch pair.

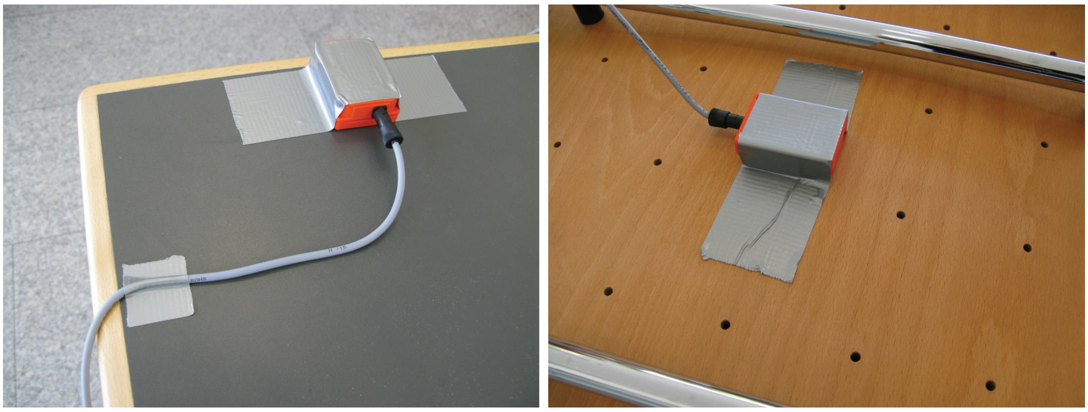


Figure 16: Xsens inertial sensors mounted on table and underneath the chair.

- Sampling frequency: 100Hz
- Connection: USB
- Client side: CRN Toolbox

#### 4.3.6 Power Sensors

**Description** Power sensors have been attached to the coffee machine and to the slicer. Power sensors allow to measure the instantaneous power used by electrical appliances. This can provide valuable information through the knowledge of the user's activity, but in previous research it was also shown that the consumption can be used to infer finer-grained information about the appliance's use, e.g. to discriminate what was cut with the slicer.

**Recording Conditions**

- Sampling frequency: 48 Hz
- Connection: USB
- Client side: CRN Toolbox

It is possible that the power sensors have been change during the experiment. The two devices have very distinct power characteristics and there it is very easy to distinguish between these two devices (also see fig. 4.3.6 ).

Figure 17: Power sensor: slicer and coffee machine power consumption were monitored.

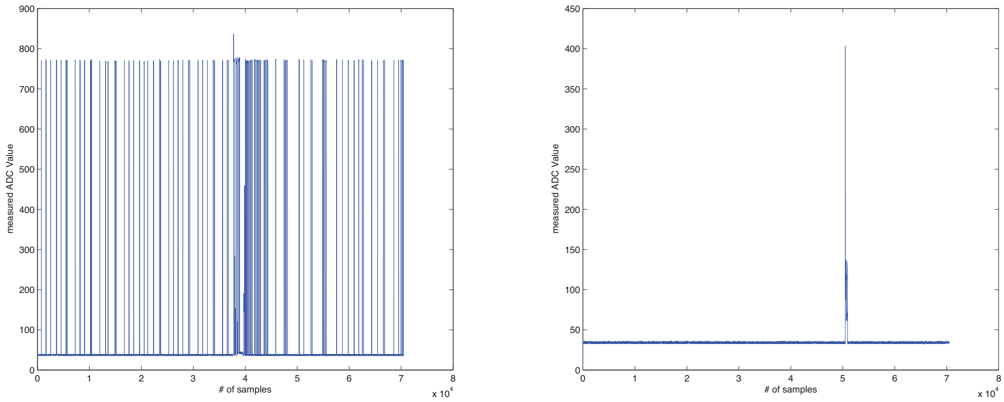



Figure 18: Left plot: power consumption of coffee machine, the machine tries to keep the water at a certain temperature and therefore a lot of power is consumed. Right plot: the slicer is only turned on when it is used, therefore the power consumption is mostly at a very low level.

### 4.3.7 Fisheye Cameras

**Description** Three LAN attached webcams have been used to collect visual information about the experiment. The AXIS camera have a fisheye shaped lense covering a large area but with an optical distortion in the field of view. The cameras are attached on the ceiling and on one side of the room, another camera was attached over the kitchen part to have a better view while the subject opens or closes the drawers. The video stream is in the ASF format with a refresh rate of 10 fps 640x480 pixels. We transcode this video stream to a mpeg4 container to reduce the file sizes without a loss of precision. The video information can be used for either activity annotation or for person tracking. In figure 20 depicted are two screenshots of pictures taken by the webcams. The three video streams are recorded with a software provided by AXIS. We noticed that there has been interferences between the ultra wide band based localization system and the webcams on network basis. The Ubisense server was in another room and the data was transmited to and from the server via Ethernet 1 Gbit connection. The switch inbetween had to transmit a lot of small multicast messages as the localization system works with this kind of messages. This reduces the available bandwidth as it is very processing intensive. We directly connected the switch to the server and thus the interferences have been reduced.

### 4.3.8 Audio Recordings

**Hardware** The audio recording can be split in two parts: The first part is a wireless audio recording of near body events. And the second part is a stationary audio recording done by four Microphones from a fixed position in the room. The on body microphones transmit the audio events near the body to the recording device. It is therefore possible to recognize activities by specific sounds. One microphone is attached near the hand, the other wireless microphone is attached to the collar.

Figure 19: An AKG condensator microphone on the left side, on the right hand side a firepod.

4 AKG C1000S MKIII microphones are positioned on tripods as ambient microphones connected via wire to the recording device. These condensator microphones have a frequency interval between 50 Hz and 20kHz. They have a kidney shaped sensitivity.

**Analog-Digital Connectivity** The microphones only provide information in an analogue form. An analogue to digital converter has to be used to transform the information. The Firepod provides this feature. 8 Channels can simultaniously be recorded with an sampling frequency up to 96kHz. During the recordings 44.1kHz was used. The synchronized and digitalized data is transfered from the firepod to the laptop using a firewire connection.

# 5 Used Sofware

## 5.1 Cubase

As described in the sensor section 4.3.8 we use the Firepod to record 6 different audio sources and time synchronize the data. On the software side Cubase allows to control the recording. The system handles the firewire data connection and stores the audio channels in files. The software gives overview of the current audio signales sent from the firewire audio device to the computer. Audio sensitivy, sampling rates and channel selections (the Firepod connects to up to 8 different audio devices) can be configured with this program.

## 5.2 Context Recognition Toolbox

The Context Recognition Network (CRN) Toolbox permits fast implementation of activity and context recognition systems. It utilizes parameterizable and reusable software components and provides a broad set of online algorithms for multi-modal sensor input, signal process- ing, and pattern recognition. It features mechanisms for distributed processing and support for mobile and wearable devices. In the data recordings we used this software as it provides drivers for different sensor systems, timestamps for data samples and online label support.

### 5.3 Labelling Tool

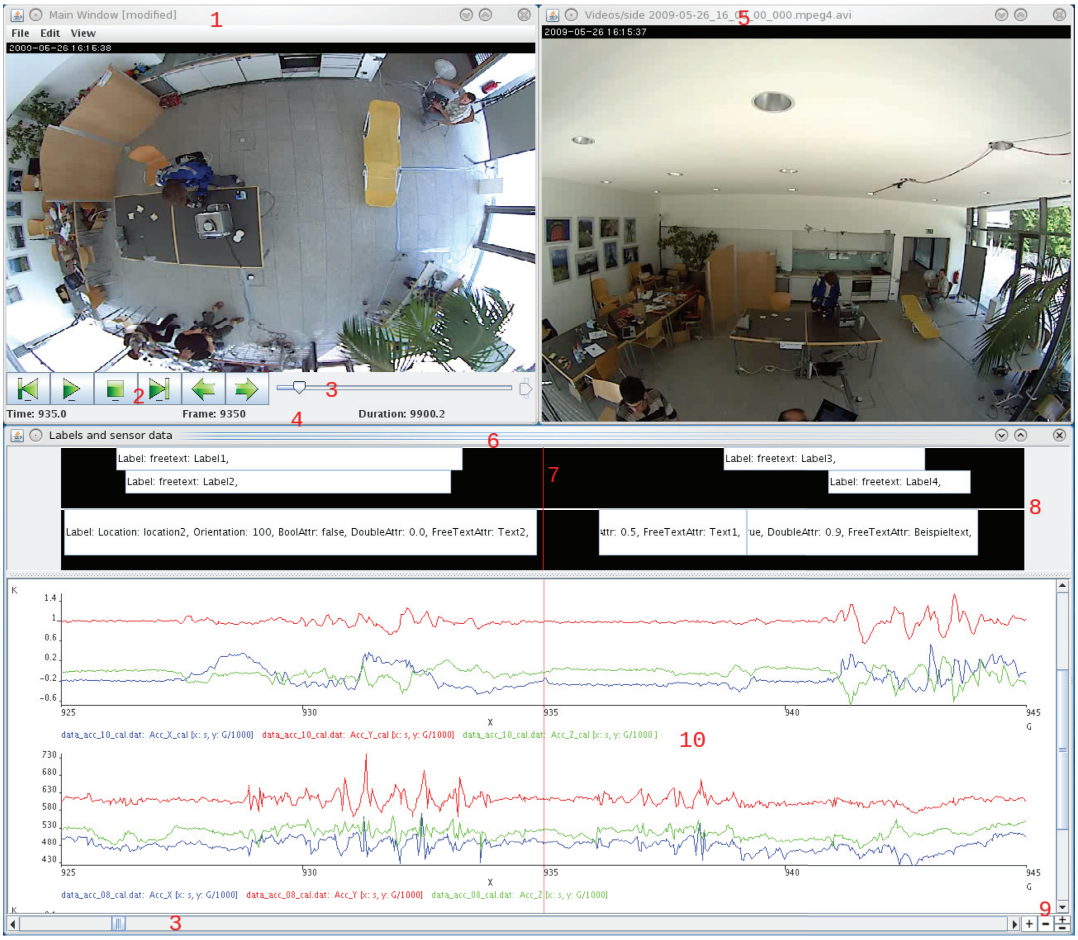



Figure 20: Screenshot of the Labelling Tool. Two video frames provide a good view of the activities. in the lower frame there the annotation part of the program is located. The datastreams can be aggregated and plotted in one plot.

As explained in earlier sections we recorded a high number of sensor video and audio data streams. We also labeled activities during the data recordings. Usually these online labels only give the rough time frame and need to be adapted offline. There are tools which allow to re-structure the labels (start and end time or activities) (for example `http://www2.ife.ee.ethz.ch/~oamft/projects/marker/index.html`) which only runs in Matlab. Other software e.g. Anvil allow to annotate video streams but cannot synchronize video and data streams. (`http://www.anvil-software.de/`). A program supporting synchronization of video and sensor data streams has not been found. We therefore implemented a Java Software providing these features.
The labelling tool runs on different OS. Video support is provided by Java Media Framework and JFFMPEG. These libraries provide synchronized players ( 3 video data streams from the webcams) and a high number of codecs. JChart2D plots the different data streams from the sensor systems. The project based tool encapsulates different sensor data streams and video files. Labels descriptions and sensor data format description are also definied in the project file. The software can also be

| | # | Min Len | Max Len | Mean Len | Tot Time |
|---|---|---|---|---|---|
| **Walk** | 628.0 | 0.3 | 146.3 | 4.8 | 3013.8 |
| **Stand** | 649.0 | 0.3 | 171.2 | 8.5 | 5511.7 |
| **Lie** | 20.0 | 4.8 | 29.7 | 18.1 | 361.7 |
| **Sit** | 71.0 | 0.8 | 274.9 | 35.7 | 2533.4 |

Table 2: Overall instances of modes of locomotion, along with minimum, maximum, average and total duration (seconds).

| | # | Min Len | Max Len | Mean Len | Tot Time |
|---|---|---|---|---|---|
| **Ambient** | 1573.0 | 0.2 | 6.3 | 1.0 | 1606.5 |
| **Objects** | 2442.0 | 0.2 | 64.5 | 2.1 | 5154.3 |

Table 3: Overall instances of hand interactions with the objects and infrastructure (seconds).

used with new sensor modalities because the data format can be specified in a format file. The program can synchronize different time dependent data streams with a master video file (Offset of different streams in relation to the video file). Therefore the synchronization gesture (clapping) is used as this gesture is easy to recognize in the video and in the data streams.
A label section in the program describes activities derived from either the movie or the data stream. An annotation has a start and end time (relative to the movie) and can have different values (e.g. combined location information and orientation information). It is also possible to have several annotation tracks to annotate different activity levels in parallel.
Depicted in figure 20 a typical screenshot. Two video frames show the kitchen scene from two different points covering most of the setup. In the lower frame there are several annotation tracks for different activity abstraction levels. The user can add at each of these tracks labels describing the activities in the videos. The data streams of several sensors are plotted in the lowest parts. Several channels can be combined to a single plot depicting semantic related sensor information.

# 6 Evaluation of the dataset acquisition

## 6.1 Evaluation of activity instances

We annotated the activity occurrences from the video footage after the recording. Currently 9 out of 60 ADL runs and 1 out of 12 drill runs are annotated.
Table 2 presents statistics on the occurrences of modes of locomotion and table 3 shows statistics on the occurrences of hand interactions with the environment and objects (an interaction is one of *reach*, *open*, *grasp*, etc). These are overall results for the 10 annotated runs. Extrapolating from this to the whole dataset, over 12000 interactions with objects and 19000 interactions with the environment may have been recorded. Tables 3 break down the hand interactions with objects in the 9 annotated ADL runs. Table 5 breaks down the right hand interactions during the drill session. As expected, activities occur roughly in multiple of 20 instances (subjects repeated 20 times the drill sequence).
By extrapolating from the currently labeled sessions, we can estimate in 25 the total number of

| | reach | move | release | stir | sip | bite | cut | spread |
|---|---|---|---|---|---|---|---|---|
| **Cup** | 23/79 | 24/130 | 18/73 | 0/3 | 4/49 | 0/0 | 0/0 | 0/0 |
| **Glass** | 12/64 | 17/103 | 11/57 | 0/1 | 1/39 | 0/0 | 0/0 | 0/0 |
| **Spoon** | 2/13 | 4/26 | 3/14 | 0/9 | 0/0 | 0/0 | 0/0 | 0/0 |
| **Sugar** | 15/16 | 18/16 | 16/12 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 |
| **Knife1** | 8/24 | 8/26 | 5/19 | 0/0 | 0/0 | 0/0 | 0/3 | 0/0 |
| **Knife2** | 9/20 | 12/34 | 9/17 | 0/0 | 0/0 | 0/0 | 0/0 | 0/7 |
| **Salami** | 23/33 | 35/39 | 20/30 | 0/0 | 0/0 | 0/0 | 0/6 | 0/0 |
| **Bottle** | 11/9 | 11/20 | 8/9 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 |
| **Plate** | 16/21 | 17/25 | 13/19 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 |
| **Cheese** | 25/25 | 37/25 | 23/19 | 0/0 | 0/0 | 0/0 | 0/1 | 0/6 |
| **Bread** | 61/62 | 99/109 | 53/52 | 0/0 | 0/0 | 25/32 | 1/2 | 0/0 |
| **Milk** | 22/33 | 19/30 | 14/21 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 |

Table 4: Number of interactions of the right/left hand with objects

| | reach | open | close | lock | unlock |
|---|---|---|---|---|---|
| **Fridge** | 43 | 20 | 22 | 0 | 0 |
| **Dishwasher** | 41 | 21 | 21 | 0 | 0 |
| **Drawer1 (top)** | 38 | 21 | 21 | 0 | 0 |
| **Drawer2 (middle)** | 32 | 21 | 21 | 0 | 0 |
| **Drawer3 (lower)** | 33 | 22 | 21 | 0 | 0 |
| **Door1** | 39 | 20 | 21 | 20 | 18 |
| **Door2** | 40 | 21 | 24 | 20 | 19 |
| **Switch** | 42 | 0 | 0 | 0 | 0 |
| **Table** | 22 | 0 | 0 | 0 | 0 |
| **Chair** | 1 | 0 | 0 | 0 | 1 |

Table 5: Number of right hand interactions with the environment (doors, drawers etc) in the drill run.

hours which were recorded and that will provide approximately 57 hours of labels. These include posture/locomotion labels which are always present and hand interactions, that often occur in overlapping fashions and including multiple objects, thus giving us a very label-intensive recording. On average, each recording run lasted 20 minutes, excluding setup and debriefing times.

## 6.2 Evaluation wireless data acquisition performance

During setup we tuned the parameters of the wireless sensors. In particular for the Bluetooth motion sensors, we started from the highest sampling rate (64Hz) and transmission of all sensor channels (nodes can locally convert raw acceleration into calibrated values), and we reduced sample rate and eliminated information that can be recovered during post-processing (table 6). Eventually, stage 3 settings correspond to the available bandwidth with DM1 ACL packets (max throughput 108.8kbps), not accounting for RFCOMM retransmissions.

| Stage (runs) | Sample rate | Data | Packet size | Total BT Byte/sec | % loss |
|---|---|---|---|---|---|
| 1(7) | 64 | 8-bit packet counter, Raw+calibrated acc., raw+amplified rate of turn | 17, 27 | 33'792 | 6.2 |
| 2(36) | 32 | (same as above) | 17, 27 | 16'896 | 8.8 |
| 3(31) | 32 | 16-bit packet counter, Raw acc., raw+amplified rate of turn | 12, 22 | 13'056 | 2.5 |

Table 6: Bluetooth motion sensor parameters during the staged link optimization. Packet sizes for body and object nodes.

Some of the objects were stored for part of the recording in the fridge or in drawers, as well as on metal shelves. Despite this unfavorable conditions and the large amount of wireless devices, the overall packet loss has been quite low after tuning. In particular, packet loss dropped to an average of 2.5% in stage 3 for the setup involving less data transmission (table 6), which we indicate as the most mature and stable. In stage 2 and part of stage 3 we systematically switched off some unused sensors during the drill sessions. For simplification reasons, we count this here as packet loss, meaning that the numbers we report are worst-case figures. This can also be seen in the left diagram in figure 21, where we can also appreciate how indeed many of the sensors were streaming with really little packet loss and only some sporadic exceptions occurred. In most of the cases the average disconnection length was below 30 seconds (right plot).

To simulate opportunistic activity recognition scenarios, it is important that more sensing modalities are present at the same time. In figure 22 (derived from the stage 3 setup) we present the time fraction in which the body and object sensors were present at the same time. The vertical axis are the stage 3 runs. The horizontal axis represents the maximum number of sensors missing at the same time. Thus, not only the overall packet loss was small, but it is also distributed in a way that, for a large part of our recordings, we had all or nearly all the sensors streaming data at the same time. For example, for run number 23 and K = 3, we see that for 95% of the time, there were at least 22 sensors running at the same time.

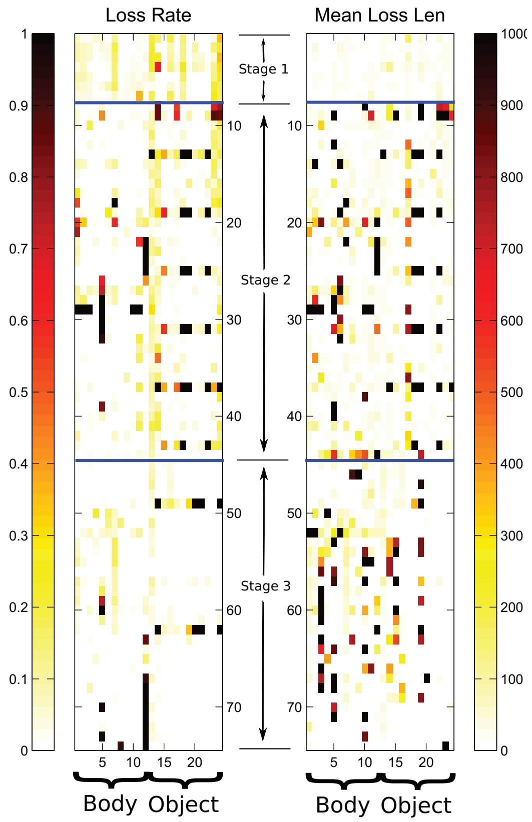



Figure 21: Loss rate (left plot) and mean disconnection length (measured in samples). Disconnection length values have been saturated to 1000 to improve the plot clarity. In the scripted runs, some sensors have been switched off because not used, this can be seen from the regular black spots on the graphs.

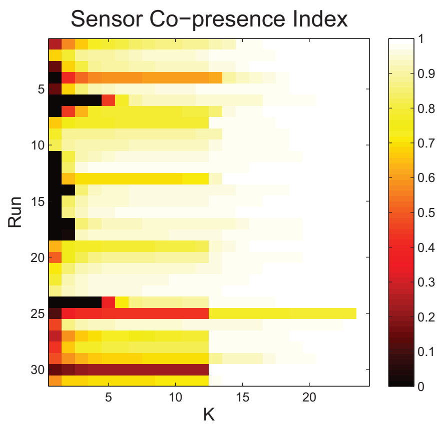



Figure 22: Fraction of the data streams where at most K sensors are missing at the same time. The leftmost column for example indicates for how much time all sensors were present all together.

# 7 Exploitation Strategies

The data set covers a high number of sensor modalities embedded in the environment, attached to objects and the subject. As presented in earlier sections activities are captured therefore from different sensors. The following steps are planned in near future:

**Standard Machine Learning Algorithms and Benchmarking** A first step to evaluate the dataset is to apply different machine learning algorithms. Training time and recognition rates are basic measures for classification algorithms.

**Unsupervised Training** Due to the high number of sensors (modalities) a training system is developed which allows to improve classification rates by attaching new sensors to the subject. Some sensor systems provide insufficient information to recognize all gestures with the same classification rate. Usually these gestures have overlapping parts in the information space. By introducing (or in our case using) extra sensor modalities this extra information helps to separate the two gestures.

**Sensor displacement** A big problem with sensors is that they have to be attached to the body. Activities and the resulting forces loosen straps and the sensors change their positions at the body. We therefore attached several acceleration sensors at different position on the body to simulate sensor displacement. Detecting displacement and reacting on the displacement will be evaluated.

**Publication of the dataset** After we evaluated the data set, the data should be made freely available in scientific community. Most test subjects agreed in this. Those who didn't will not be publicated.

# 8 Conclusion

We motivated in this document the need for the huge and complex data recording. The used and presented sensor systems attached to objects, embedded in the environment and fixed to the body capture activities in different ways and from different perspectives. Tools for video annotation of the activities have been developed after the data recordings and we currently annotating activities and assign the activities to different activity classes and levels. The data set is the fundament for the development of opportunistic algorithms and will be used in this project for benchmarking of well known standard algorithms and algorithms developed in this project. Due to the fact that there are a high number of different sensor modalities capturing a high number of different gestures this data set will be very valuable for the scientific community. We already got requests for accessing the data set from different groups.

## References

[1] Marc Bächlin, Daniel Roggen, and Gerhard Tröster. Context-aware platform for long-term life style management and medical signal analysis. In *Proceedings of the 2nd SENSATION International Conference*, 0 2007.

[2] Gerald Bauer, Karl Stockinger, and Paul Lukowicz. Recognizing the use-mode of kitchen appliances from their current consumption. In *Smart Sensing and Context, EUROSSC09*, pages 163–176, 2009.

[3] Stephen Intille, Kent Larson, Emmanuel Tapia, Jennifer Beaudin, Pallavi Kaushik, Jason Nawyn, and Randy Rockinson. Using a live-in laboratory for ubiquitous computing research. pages 349–365. 2006.

[4] Matthias Kranz, Albrecht Schmidt, Alexis Maldonado, Radu Bogdan Rusu, Michael Beetz, Benedikt Hoernler, and Gerhard Rigoll. Context-aware kitchen utilities. In *TEI '07: Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 213–214, New York, NY, USA, 2007. ACM Press.

[5] Gerald Pirkl, Karl Stockinger, Kai Kunze, and Paul Lukowicz. Adapting magnetic resonant coupling based relative positioning technology for wearable activitiy recogniton. *Twelfth IEEE International Symposium on Wearable Computers (ISWC 2008)*, pages 47 – 54, 2007.

[6] Thomas Stiefmeier, Daniel Roggen, Georg Ogris, Paul Lukowicz, and Gerhard Tröster. Wearable activity tracking in car manufacturing. *IEEE Pervasive Computing*, 7(2):42–50, 2008.

[7] Moritz Tenorth, Jan Bandouch, and Michael Beetz. The TUM kitchen data set of everyday manipulation activities for motion tracking and action recognition. In *IEEE Int. Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS). In conjunction with ICCV2009*, 2009.

[8] Moritz Tenorth, Jan Bandouch, and Michael Beetz. The TUM kitchen data set of everyday manipulation activities for motion tracking and action recognition. In *IEEE Int. Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS). In conjunction with ICCV2009*, 2009.

[9] Tim van Kasteren, Athanasios Noulas, Gwenn Englebienne, and Ben Kröse. Accurate activity recognition in a home setting. In *UbiComp '08: Proceedings of the 10th international conference on Ubiquitous computing*, pages 1–9, New York, NY, USA, 2008. ACM.

[10] D. H. Wilson and C. Atkeson. Simultaneous tracking and activity recognition (star) using many anonymous, binary sensors. In *Pervasive Computing*, pages 62–79, 2005.

[11] Jianxin Wu, A. Osuntogun, T. Choudhury, M. Philipose, and J.M. Rehg. A scalable approach to activity recognition based on object use. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, Oct. 2007.

# A Sensor Data File Format

## A.1 Motion Jacket Data format

Each file, named `data_jacket.dat`, contains one row per sample and each row has 67 columns, including two timestamps and five groups of 13. The meaning of the groups is summarized in Table 7, where the symbols stand for:

- T(s) = Time in seconds
- T(us) = Time in microseconds
- BAC = Mid back sensor (reference)
- RUA = Right upper arm sensor
- RLA = Right lower arm sensor
- LUA = Left upper arm sensor
- LLA = Left lower arm sensor

Each sensor has 13 data columns, whose meaning is illustrated in Table 17, where the symbols stand for:

- Ax, Ay, Az = Acceleration signal for axes x, y and z
- Gx, Gy, Gz = Gyroscope signal for axes x, y and z
- Mx, My, Mz = Magnetic field signal for axes x, y and z
- Q1-4 = Quaternions indicating orientation

| Col # | 1 | 2 | 3-15 | 16-28 | 29-41 | 42-54 | 55-67 |
|---|---|---|---|---|---|---|---|
| Descr. | T(s) | T(us) | BAC | RUA | RLA | LUA | LLA |

Table 7:

| Col # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Descr. | Ax | Ay | Az | Gx | Gy | Gz | Mx | My | Mz | Q1 | Q2 | Q3 | Q4 |

Table 8:

| Col # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Descr. | T(s) | T(us) | Sn | $ADC_1$ | $ADC_2$ | $ADC_3$ | $ADC_4$ | $ADC_5$ | $ADC_6$ | $ADC_7$ | $ADC_8$ | $ADC_9$ |
| Col # | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | | | |
| Descr. | $AMP_1$ | $AMP_2$ | $AMP_3$ | $AMP_4$ | $AMP_5$ | $AMP_6$ | $AMP_7$ | $AMP_8$ | $AMP_9$ | | | |

Table 9: Log file format of magnetic coupling based sensor

## A.2 Magnetic Sensor Data Format

Data from the magnetic sensors are saved in the file `data_magnetic.dat`. Each line of the log file has 12 columns, as described in Table 9.
The symbols have the following meanings:

- T(s) = Time in seconds
- T(us) = Time in microseconds
- Sn = Sequence number
- ADC1-9 = Raw ADC values measured at the receiver coils, $ADC_1$ = ADC value at x receiver coil while transmitter coil x transmitted, $ADC_2$ = ADC value at y receiver coil while transmitter coil x transmitted, $ADC_3$ = z receiver coil while transmitter coil x transmitted, $ADC_4$ = x receiver coil while transmitter coil y transmitted, ... $ADC_9$ = ADC value at z receiver coil while transmitter coil z transmitted
- AMP1-9 = $AMP_i$ Amplification Levels of corresponding $ADC_i$ channel,

## A.3 Bluetooth Acceleration Sensors Data Format

Each file, named `data_acc_xx_cal_filled.dat.dat`, contains one row per sample and the meaning of the columns is explained in Table 14. The symbols stand for:

- T(s) = Time in seconds
- T(us) = Time in microseconds
- ID = Sensor ID
- Pkt # = Packet number (8 bits for sessions 1-7, 16 bits for sessions 8-12)
- Axu, Ayu, Azu = Uncalibrated acceleration signals for axes x, y and z
- Ax, Ay, Az = Calibrated acceleration signals for axes x, y and z
- Gx, Gy = Gyroscope signals for axes x and y
- Gxa, Gya = Amplified gyroscope signals for axes x and y
- G_ref = Gyroscope reference

`xx` in filename are the ID numbers of the sensors, as listed in Table 11.

| Col # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Descr. | T(s) | T(us) | ID | Pkt # | Axu | Ayu | Azu | Ax | Ay | Az | Gx | Gxa | Gy | Gya | Gref |

Table 10:

| Position | ID |
|---|---|
| Left upper arm up | 10 |
| Left upper arm down | 26 |
| Left wrist | 27 |
| Left hand | 21 |
| Right upper arm up | 25 |
| Right upper arm down | 11 |
| Right wrist | 24 |
| Right hand | 28 |
| Right knee up | 5 |
| Right knee down | 23 |
| Right hip | 8 |
| Back | 22 |

Table 11:

## A.4 Sun SPOT Data format

Each file, named `data_shoeankle_{left|right}.dat`, contains one row per sample, and the meaning of the columns is explained in Table 12. The symbols stand for:

- Tclient(ms) = Host computer time in milliseconds
- Ttracker(ms) = Sun SPOT time in milliseconds
- AccX, AccY, AccZ = Acceleration signals for axes x, y and z in $m/s^2$.
- AccTotal = Total acceleration in $m/s^2$, computed from AccX, AccY, and AccZ.
- TiltX, TiltY, TiltZ = Inclination of the axes x, y and z in radians with respect to the total acceleration.

As for the use of the Sun SPOT acceleration data, there are some issues to notice. First, no packet counter has been used, but a timestamp has been added to each sample by the Sun SPOT

| Col # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Descr. | Tclient(ms) | Ttracker(ms) | AccTotal | AccX | AccY | AccZ | TiltX | TiltY | TiltZ |

Table 12: Sun SPOT acceleration sensor file format.

(referred to as tracker in the dataset) instead. However, this may make it difficult to recognize whether samples have been lost in the wireless transmission or not, as there was a high variability of the intervals between two acceleration measurements (i.e. between the tracker time stamps). Also, in many sessions, at least one of the Sun SPOTs stopped transmitting data during the session due to an empty battery. In order to cope with this irregularities, a special processing will be needed, where two complementary methods are proposed:

- Identification of data segments with missing measurements, for which purpose a sliding window with a size of e.g. 1000 ms could be used. Whenever the number of measurements within the window drops below a certain threshold which can be derived from the known sampling rate, the segment will be discarded.
- Linear interpolation between the measurements at two successive points in time, which is possible as the tracker timestamps have been recorded right before acquiring the acceleration data.

## A.5 Inertiacubes Data format

Each file, named `data_shoetoebox_{left|right}_x.dat`, contains one row per sample and the meaning of the columns is explained in Table 13. The symbols stand for:

- T(s) = Time in seconds
- T(us) = Time in microseconds
- eulerX, eulerY, eulerZ = (yaw, pitch, roll) global Euler angles ($deg$).
- navAccX, navAccY, navAccZ = Acceleration in the navigation coordinate frame. Best estimate on global accelerometer measurements with calibration, and current sensor orientation applied, and gravity subtracted ($m/s^2$).
- bodyAccX, bodyAccY, bodyAccZ = Acceleration in sensor body coordinate frame. Gravity component is not removed ($m/s^2$).
- angVelBodyX, angVelBodyY,angVelBodyZ = Angular rotation speed in body coordinate frame. This is the processed angular rate, with current biases removed. This is the angular rate used to produce orientation updates ($rad/s$).
- angVelNavX, angVelNavX, angVelNavZ = Angular rotation speed in world coordinate frame, with boresight and other transformations applied ($rad/sec$).
- compass (mh) = Magnetometer heading, computed based on current orientation ($deg$).

## A.6 Bluetooth Acceleration+Gyroscope Sensors Data Format

Each file, named `data_acc_bt_xxx_object_cal_filled.dat`, contains one row per sample and the meaning of the columns is explained in Table 14. The symbols stand for:

- T(s) = Time in seconds
- T(us) = Time in microseconds

| Col # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Descr. | T(s) | T(us) | eX | eY | eZ | nAccX | nAccY | nAccZ | bAccX | bAccY | bAccZ |
| Col # | 15 | 16 | 17 | 18 | 19 | 20 | 21 | | | | |
| Descr. | bAVX | bAVY | bAVZ | nAVX | nAVY | nAVZ | mh | | | | |

Table 13: Shoe toe box Inertia Cube 3 file format.

| Col # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Descr. | T(s) | T(us) | ID | Pkt # | Axu | Ayu | Azu | Ax | Ay | Az | Gx | Gxa | Gy | Gya | Gref |

Table 14:

- ID = Sensor ID
- Pkt # = Packet number (8 bits for sessions 1-7, 16 bits for sessions 8-12)
- Axu, Ayu, Azu = Uncalibrated acceleration signals for axes x, y and z
- Ax, Ay, Az = Calibrated acceleration signals for axes x, y and z
- Gx, Gy = Gyroscope signals for axes x and y
- Gxa, Gya = Amplified gyroscope signals for axes x and y
- G_ref = Gyroscope reference

`xxx` in filename are the ID numbers of the sensors, as listed in Table 15.

## A.7 USB Acceleration Sensors Data Format

## A.8 Ubisense Data Format

The logfile **data_ubisense.log** holds all positions of each tag - the tags are mixed in the logfile. There are 7 columns in the file.

- T(s) Time stamp in s
- T(us)=Time in microseconds
- tag id = id of the Tag of which the position has been estimated
- x position of the tag in mm
- y position of the tag in mm
- z position of the tag in mm
- acc = accuracy of the measurement - if it is high - it is very accurate.

| Object | ID |
|---|---|
| Salami | 112 |
| Water | 113 |
| Cheese | 114 |
| Milk | 118 |
| Cup | 111 |
| Spoon | 119 |
| Knife | 123 |
| Plate | 124 |
| Bread | 116 |
| Sugar | 120 |
| Knife for salami | 117 |
| Glass | 125 |

Table 15:

Table 16: default

| Column # | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Descr. | T(s) | T(us) | tag id | x pos mm | y pos mm | z pos mm | acc |

## A.9 Reed Switches Data format

Each file, named `data_reed.dat`, contains one row per sample and each row has 18 columns. The meaning of the columns is summarized in Table **??**, where the symbols stand for:

- T(s) = Time in seconds
- T(us) = Time in microseconds
- F1-3 = Fridge door switches
- W1-3 = Dishwasher door switches
- U = Upper drawer switch
- M1-3 = Middle drawer switches
- L1-3 = Lower drawer switches

## A.10 Force Resistive Sensors Data Format

## A.11 Table and Chair Xsens Data format

The two sets of sensor data are stored in files named `data_xchair.log` and `data_xtable.log`. Each file contains one row per sample and each row has 11 columns, including two timestamps and

| Col # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11-13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Descr. | T(s) | T(us) | W1 | F3 | F2 | F1 | M1 | M2 | M3 | L3 | 1 | L2 | U | W3 | L1 | W2 |

Table 17:

| Col # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Descr. | T(s) | T(us) | Ax | Ay | Az | Gx | Gy | Gz | Mx | My | Mz | Q1 | Q2 | Q3 | Q4 |

Table 18:

9 sensor measurements. The column content is summarized in Table 18, where the symbols stand for:

- T(s) = Time in seconds
- T(us) = Time in microseconds
- Ax, Ay, Az = Acceleration signal for axes x, y and z
- Gx, Gy, Gz = Gyroscope signal for axes x, y and z
- Mx, My, Mz = Magnetic field signal for axes x, y and z

## A.12 Power Sensor Data Format

The two power sensor data are recorded in separate files (`data_power1.log` and `data_power2.log`). Each log file has 4 columns, containing time stamps and power measurement (see Table 19).

| Col # | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Descr. | T(s) | T(us) | ID | ADC |

Table 19:

The symbols stand for:

- T(s) = Time in seconds
- T(us) = Time in microseconds
- ID = Sensor ID
- ADC = ADC reading, proportional to power consumption

Table A.12 shows the mapping between IDs and devices.

## A.13 Fisheye Cameras Data Format

| Device | ID |
|---|---|
| Slicer | 1 |
| Coffeemachine | 2 |