Collecting complex activity datasets in highly rich networked sensor environments

Daniel Roggen, Alberto Calatroni, Mirco Rossi
Thomas Holleczek, Kilian Förster, Gerhard Tröster
Wearable Computing Laboratory, ETH Zürich
Email: daniel.roggen@ife.ee.ethz.ch

Paul Lukowicz, David Bannach, Gerald Pirkl
Embedded Systems Laboratory
University of Passau
Email: name.surname@uni-passau.de

Alois Ferscha, Jakob Doppler
Clemens Holzmann, Marc Kurz, Gerald Holl
Institute for Pervasive Computing
Johannes Kepler University Linz
Email: ferscha@soft.uni-linz.ac.at

Ricardo Chavarriaga, Hesam Sagha, Hamidreza Bayati,
Marco Creatura[+], José del R. Millàn
Defitech Foundation Chair in Non-Invasive Brain-Machine
Interface, Ecole Polytechnique Fédérale de Lausanne
Email: name.surname@epfl.ch
[+]University of Genova, Department of Informatics,
Systems and Telematics, Italy

Abstract—We deployed 72 sensors of 10 modalities in 15 wireless and wired networked sensor systems in the environment, in objects, and on the body to create a sensor-rich environment for the machine recognition of human activities. We acquired data from 12 subjects performing morning activities, yielding over 25 hours of sensor data. We report the number of activity occurrences observed during post-processing, and estimate that over 13000 and 14000 object and environment interactions occurred. We describe the networked sensor setup and the methodology for data acquisition, synchronization and curation. We report on the challenges and outline lessons learned and best practice for similar large scale deployments of heterogeneous networked sensor systems. We evaluate data acquisition quality for on-body and object integrated wireless sensors; there is less than 2.5% packet loss after tuning. We outline our use of the dataset to develop new sensor network self-organization principles and machine learning techniques for activity recognition in opportunistic sensor configurations. Eventually this dataset will be made public.

I. Introduction

Sensor networks (integrated into objects, on body or in the environment) allow to sense the physical world and persons acting in it [1], [2]. In wearable and pervasive computing, this allows to detect the user's *context* and provide ambient intelligence environments or smart assistance when and where users need it, proactively with minimal interaction. Human activities and gestures are important aspects of context. Applications include gesture-based human-computer interaction, healthcare [3], or industrial workers [4], and is key to many other intelligent environments [5].

A. Problem statement

The prevailing activity recognition approach is to deploy application-specific sensors at well defined locations. This is often not desirable or tedious. Users are at times in highly instrumented environments and at other times in places with little sensor infrastructure. Ensuring that sensors are placed on-body at precise locations day in, day out is cumbersome. Users

already (or soon will) carry many sensor enabled devices, such as mobile phones (e.g. with GPS and motion sensors), headsets, or intelligent motion-sensing garments. As the user changes location, leaves devices behind, picks up new ones and changes her outfit, the sensing environment changes.

We envision activity recognition from *opportunistically* discovered sensors (*opportunistic sensor configurations*). A number of approaches exist for the coordinated emergence of sensing networks [6], [7]. However, interpreting the sensor data remains a challenge, as there is no a-priori knowledge about number, kind or placement of discovered sensors. Thus, the mapping between sensor signals and activities cannot be learned at design time. Within the EU project OPPORTUNITY we investigate how to address these challenges [8].

In order to characterize these methods, empirical validation is necessary against a reference baseline. Thus, we set out to acquire a large scale multimodal data set of naturalistic human activities in a sensor rich environment. Various combinations of opportunistic methods and/or available sensors (simulating opportunistic sensor configurations) can then be benchmarked.

B. Paper contribution

We report on the acquisition of a dataset of naturalistic human activities in a sensor rich environment: a room simulating a studio flat with kitchen, deckchair, and outdoor access where subjects performed daily morning activities. 15 networked sensor systems were deployed, with 72 sensors of 10 modalities, integrated in the environment, in objects, and on the body. It is an example of the deployment of a large number of networked sensor systems of different origins (proprietary and custom, from different manufacturers or universities) for the application domain of activity recognition.

The main contributions are:

- A dataset of complex, interleaved and hierarchical naturalistic activities, with a particularly large number of

atomic activities (more than 27'000), collected in a very rich sensor environment, compared to other datasets. This makes this dataset well suited to benchmark various activity recognition approaches, and to investigate e.g. multimodal data fusion, reasoning, or activity and scenario modeling. This dataset will eventually be publicly available with reference to the original paper (this article).

- A description of the approach, lessons learned, and best practices for the deployment and data acquisition from similar complex deployments of heterogeneous networked sensor systems, with emphasis on the sensor environment setup, and data management process (e.g. synchronization, curation, distribution).

We present a first data quality evaluation for 12 on-body and 12 object integrated wireless motion sensors and we report on the number of activity occurrences.

II. STATE OF THE ART

A. Activity recognition in multimodal sensor networks

Many sensors can be used for activity recognition (table I). Sensor networks enabling node mobility allow to use a combination of on-body, object, and ambient sensors. Multiple sensors are usually beneficial: i) some modalities or sensor placement may be more suitable for some activities [4]; ii) fusing the decision of multiple sensor specific classifiers may outperform a single classifier [9]. Usually machine learning is used to interpret sensed data into activities. During a training phase, the user(s) performs multiple times the activities of interest. Then, toolboxes such as WEKA[1] may be used to train classifiers to recognize the signal templates of interest. Network-oriented data processing toolboxes such as the CRN Toolbox [10], TITAN [11] or SPINE [12] can then execute the recognition algorithms.

B. Datasets for activity recognition

A few of the more known datasets are: the PlaceLab dataset, focusing on ambient and object sensing [13]; Van Kasteren's dataset [14] with particularly long recordings (month-long) but with fewer sensors, and the Darmstadt routine dataset used for unsupervised activity pattern discovery [15], that is a long recording from body activity collected by the Porcupine system [16]. The TUM Kitchen data set was recorded for video-based activity recognition [17]. It also contains RFID and reed switch data, but it does not include on-body sensors. Most of the existing datasets are not sufficiently rich to investigate opportunistic activity recognition, where a high number of sensors is required on the body, in objects and in the environment, with a high number of activity instances.

C. Context recognition in opportunistic sensor configurations

Within the EU FP7 FET-Open project OPPORTUNITY, we develop mobile systems to recognize human activity in opportunistic sensor setups [8]. We envision developments along self-organized sensing, opportunistic context recognition methods and autonomous adaptation (see figure 1). We summarize a few results below (further details in [8], especially with respect to sensor self-organization).
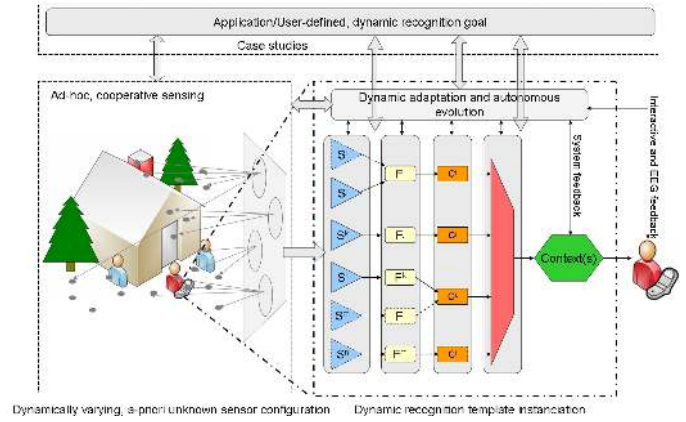
Fig. 1. The OPPORTUNITY system: the user's mobile device triggers the sensor nodes self-organization. Each sensor node (a Context Cell) is an autonomous unit capable of self-description and other self-* properties that infers the user's context from the sensor data. It can update its probabilistic context representation (online learning) from neighbors' inputs, share it with the mobile device, and update it's self-description, thus forming an autonomously evolving and adapting sensor ecology.

In opportunistic activity recognition, there is not necessarily a static signal pattern to activity mapping. Thus, classification methods must be robust to possible signal variations. We showed how activity recognition can be made resilient to small changes in on-body sensor placement using unsupervised techniques [18], principles of body mechanics [19], or evolutionary techniques [20]. We showed that sensors can autonomously recognize their on-body position [21] and their symbolic location in the environment [22]. We showed principles that allows one sensor node to autonomously learn how to recognize user activity from another one, thus allowing an activity recognition system to autonomously expand to new resources discovered or introduced in the environment, without

Where	Sensors	Observation
EBO	Microphone	Speaker recognition, localization by ambient sounds, activity detection, object self-localization
EBO	Accelerometers or gyroscopes	Body movement patterns, object use, ambient infrastructure
–BO	Magnetometer	Orientation of the body or objects
–BO	Inertial sensor (acc, rot, mag)	Absolute orientation, multiple sensors for body model reconstruction
–B–	Relative magnetic sensing	Position of body parts w.r.t. a reference
E––	Camera	Localization, body model reconstruction
E–O	Reed switches	Use of objects, ambient infrastructure
E––	UWB localization	User localization
E–O	RFID	Use of objects, ambient infrastructure
E––	Proximity infrared	Movement detection, localization
–B–	Pressure	Vertical motion in elevator or staircase
–B–	Light sensor (visible, IR, UV)	Localization of windows, lamps, light tubes, sunshine
–B–	Skin temperature	Health state (e.g. fever)
E––	Environment temperature	Outdoor, indoor
–B–	Humidity	Physical activity
–B–	Strain, stress	User's breathing (respiration belt), movement (strain sensors in clothes)
–B–	ECG	Physical activity, health state
–B–	EMG, EOG	Muscle (EMG) and eye (EOG) activation
–B–	EEG, fNIR	Cognitive states

TABLE I

COMMON SENSOR MODALITIES FOR ACTIVITY RECOGNITION. 'WHERE' INDICATES BODY (B), OBJECTS (O), OR ENVIRONMENT (E) SENSORS. THE SENSORS IN THE FIRST HALF OF THE TABLE WERE USED HERE.
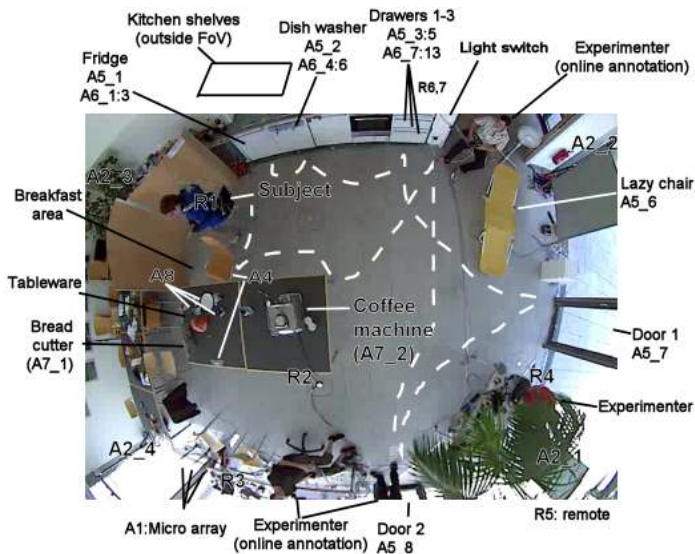
Fig. 2. View of the room from top. Dashed line: typical user trajectory in the drill run. In the ADL runs subjects moved with extreme variability.
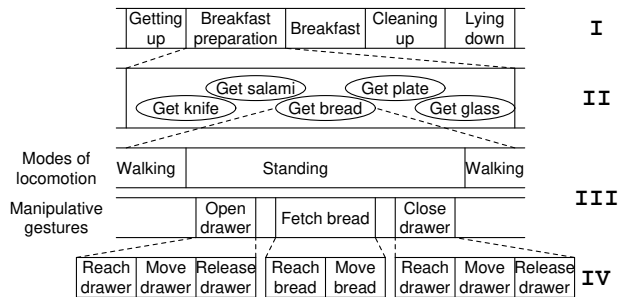

Fig. 3. Temporal decomposition of activities. Level I is the highest activity level available in the setup. Level II zooms in into one high level activity, in this level the activities are not temporal ordered and depend on the execution sequence of the subject. Logical, physiological and spatial limitations distinguish the order of activities 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.

user intervention [23]. We showed that adaptive methods can lead to an autonomous system capable of self-improvement, by using minimalist or even unconscious user feedback [24].

III. THE OPPORTUNITY DATASET SCENARIO

We designed the activity recognition environment and scenario to generate many activity primitives, yet in a realistic manner. We purposely did not record human behavior in daily life. Other datasets exist for this purpose, and the need for a highly multimodal setup is impractical for use over weeks, and may lead to privacy concerns. Instead, our focus was to maximize the number of activity instances that were collected, while keeping their execution naturalistic. We achieved this by relying on a high-level script and leaving free interpretation to the users, and even encouraging them to perform as naturally as possible with all the variations they were used to. Subjects operated in a room simulating a studio flat with a deckchair, a kitchen, doors giving access to the outside, a coffee machine, a table and a chair (figure 2).

In order to simulate opportunistic sensor configurations, the environment must be sensor-rich: i) all activities should be sensed by multiple sensors; ii) multiple sensors in close proximity allow to study robustness against sensor placement variability; iii) sensors of different modalities but sensing information related to a common activity allow to study the dynamic replacement of one modality by another; iv) multiple sensors of identical modalities but from different systems allow to assess the effects of calibration, resolution, or sample rate variations. Thus we deployed multiple wireless and wired networked sensor systems from different origins in close proximity. This leads to a challenging sensor setup with respect to data acquisition, synchronization, and curation.

A. Scenario script

Each subject performed 5 times an activity of daily living (ADL) 'run' and one 'drill run' designed to generate a large number of activity instances. The ADL run consists of temporally unfolding situations. In each situation (e.g. *preparing*

sandwich), composite activities (e.g. *cutting bread*) occur as well as atomic activities (e.g. *reach* for bread, *move* to bread cutter, *operate* bread cutter). This allows to look at activity recognition at various abstraction levels. A video presentation of the dataset is available at http://vimeo.com/8704668.

1) ADL run: ADL runs consist of this activity sequence:
1) Start: lying on the deckchair, get up
2) Groom: move in the room, check that all the objects are in the right places in the drawers and on shelves
3) Relax: go outside and have a walk around the building
4) Prepare coffee: prepare a coffee with milk and sugar using the coffee machine
5) Drink coffee: take coffee sips, act naturally in the environment
6) Prepare sandwich: include bread, cheese and salami, using the bread cutter and various knifes and plates
7) Eat sandwich
8) Cleanup: put objects used to original place or dish washer, cleanup the table
9) Break: lie on the deckchair

On a higher abstraction level, this sequence may be summarized as 'get up', 'coffee', 'sandwich', 'clean' and 'break'. On a finer level, a large number of gesture primitives can be observed (see figure 3).

2) Drill run: Subjects performed 20 repetitions of the following sequence to generate many activity instances:
1) Open and close the fridge
2) Open and close the dishwasher
3) Open and close 3 drawers (at different heights)
4) Open and close door 1
5) Open and close door 2
6) Turn on and off the lights
7) Clean table
8) Drink (standing)
9) Drink (sitting)

B. Sensor systems

Multiple sensor systems (table II) were deployed[2] on body (fig. 4), on objects (fig. 5) and in the environment (fig. 2). Wireless sensors included 24 custom Bluetooth wireless acceleration and gyroscope sensors, 2 Sun SPOTs (802.15.4) and 2 Intertiacube3 (custom 2.4GHz protocol). In addition, the

---

[2]A sensor system is a set of nodes forming a sensor network, together with the corresponding software to acquire their data.
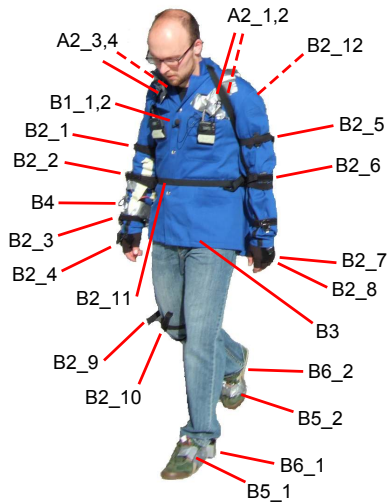
Fig. 4. Location of the body-worn sensors on the subject.

Fig. 5. The objects are instrumented with an acceleration and rate of turn sensor node. Visible are also the bread slicer (right), an XSense measuring table vibration (far right), and pressure sensors under some of the tableware. Not visible: the identically instrumented bread and salami.

Ubisense localization system operates in the 5 GHZ frequency band. The custom magnetic field sensor emits an EM field for relative positioning. Close proximity between systems makes this scenario challenging for the wireless nodes.

Data acquisition at a single point is challenging in an heterogeneous system. Commercial systems usually have proprietary sensor network management softwares, that are difficult to integrate in a larger framework; our custom systems were generally easier to integrate. Seven computers acquired the data from specific sensor systems (table III). On-body sensors were managed by a dedicated laptop in a backpack (local storage as there is no WLAN outside of the room). Ambient and object sensors were acquired by multiple computers according to the bandwidth required (e.g. video and audio streams each on a dedicated computer), the cabling possibilities (e.g. to deploy an antenna), the distance to the wired ambient sensors, and the need to minimize supplemental wireless transmissions to minimize the risk of data loss.

C. Experimental protocol

The subject was instructed with the overall experimental protocol. She then executed the 5 ADL runs, with a 10-20 minutes break between runs to copy data, check battery levels

and ensure correct system behavior. An instructor followed the subject in the first run to indicate her the sequence of activities. The subject acted alone in the following runs. A run lasted 15-25 minutes. We placed little constraints in the way users should perform during the runs. We instructed them to follow the high-level action sequence (from getting up to preparing the coffee, preparing a sandwich, etc) and to perform naturally. Users were allowed to interleave their actions (e.g. start the sandwich preparation while still taking sips from the coffee cup). Later, the subject executed the drill run (20-35 minutes). Also here we encouraged users to perform naturally (e.g. we told subjects not to hesitate to use different hands when interacting with the environment/objects).

Batteries were regularly recharged or exchanged during breaks if their operation time was too short (e.g. the Motion Jacket batteries lasted about 2 hours, the Bluetooth sensors operated for a full day). Room lighting (fluorescent tubes) were always on to minimize differences due to external lighting, and the blinds of the room side exposed to the sun were closed.

ID	Sensor system	Location and observation
B1	Commercial wireless microphones	Chest and dominant wrist. Senses user activity
B2	Custom wireless Bluetooth acceleration sensors [25]	12 locations on the body. Senses limb movement
B3	Custom motion jacket [4]	Jacket including 5 commercial RS485-networked XSense inertial measurement units
B4	Custom magnetic relative positioning sensor [26]	Emitter on shoulder, receiver on dominant wrist. Senses distance of hand to body
B5	Commercial InertiaCube3 inertial sensor system	One per foot, on the shoe toe box. Senses modes of locomotion
B6	Commercial Sun SPOT acceleration sensors	One per foot, right below the outer ankle. Senses modes of locomotion
O1	Custom wireless Bluetooth acceleration and rate of turn sensors	On 12 objects used in the scenario. Senses object use
A1	Commercial wired microphone array	4 at one room side. Senses ambient sound
A2	Commercial Ubisense localization system	Corners of the room. Senses user location
A3	Axis network cameras	3 locations, for localization, documentation and visual annotation
A4	XSense inertial sensor [4]	On the table and chair. Senses vibration and use
A5	USB networked acceleration sensors [27]	8, on doors, drawers, shelves and lazy chair. Sense usage
A6	Reed switches	13, on doors, drawers, shelves. Sense usage, provides ground truth
A7	Custom power sensors	Connected to coffee machine and bread cutter. Senses usage
A8	Custom pressure sensors	3 on the table, user placed plates and cups on them. Senses usage

TABLE II
SENSOR SYSTEMS DEPLOYED IN THE EXPERIMENTAL SETUP.

ID	Records sensor systems	Nature and location	Data acquisition
R1	B2, B3, B4	Laptop, on body in a backpack	CRN Toolbox [10]
R2	A2, A4, A7	Desktop PC	CRN Toolbox
R3	B1, A1	Laptop (static)	Audio acq. software
R4	B5, B6	Laptop (carried by experimenter, following subject)	Commercial proprietary software
R5	A3	Laptop (static)	Axis proprietary
R6	A5	Laptop (static)	Dedicated software
R7	O1, A6	Laptop (static)	CRN Toolbox

TABLE III
DATA ACQUISITION INFRASTRUCTURE AND SOFTWARE.

IV. DATA MANAGEMENT

Following data acquisition, data must be prepared in an adequate form to be analyzed (data curation).

A. Dataset curation repository

Multiple partners access the dataset for post-processing. The large amount of raw data (>130GB) and the need for read/write access led us to store the raw data on a backuped system accessed using the synchronization tool `unison`[3]. `Unison` allows efficient bidirectional automatic synchronization (i.e. local modified data are sent to the server, and server modified data are sent locally) while minimizing network load (only missing data is transferred). `Unison` ensures no data is lost if multiple synchronization are done at the same time. Once data curation is completed this repository will become read-only. Small frequently modified working files (e.g. dataset annotation, documentation) are shared via the `subversion` version control system[4] that keeps a history of changes.

B. Synchronization of data streams

Ideally, the nodes of a sensor network are synchronized and data samples are flagged with the acquisition time to simplify the reconstruction of a single synchronized data stream, despite variable network delays. In heterogeneous systems, the following problems arise: i) sensors may not flag data with a timestamp during acquisition, ii) sensors may be proprietary and cannot be modified to support timestamps; iii) even with timestamps, the synchronization across sensor network boundaries remains an issue.

Our approach is the following: i) we synchronize the data from the sensor systems offline; ii) data are flagged with the time of reception. In activity recognition synchronization requirements (< 100ms) are defined by human dynamics. All sensors on a given recording computer (table III) share the same clock domain and can easily be synchronized. In wireless networks, variable delays are common, e.g. due to wireless retransmission in case of errors or the burst transmission of data to make best use of the air interface. They translate into irregular time intervals between received data packets. However, all our sensors guarantee a regular sample rate and transmit a data packet counter. Thus, we used a least-square regression (LSQR) to compute a linear fit between data packet number and reception time. We then redefined the time of reception of the packets to ensure regular time interval between packets. Over 10+ minutes of recording, the error in the LSQR becomes negligible. Since data packets may be missing we rely on the packet counter to compute the regression correctly despite data loss (this approach is also discussed in [28]).

During postprocessing, we determined the time offset between the recording computers by signal inspection. For instance, on-body acceleration is referenced to the video time; then object acceleration is referenced to the video time (i.e. when the user grasps an object there is a corresponding acceleration signal); and eventually, video, body and object acceleration are referred to the same time.

[3] http://www.cis.upenn.edu/ bcpierce/unison/
[4] subversion.tigris.org

	#	Min Len	Max Len	Mean Len	Tot Time
Walk	1414	0.3	242.6	5.6	7900.5
Stand	1043	0.2	171.2	7.5	7770.9
Lie	56	0.9	166.6	21.8	1219.5
Sit	127	0.8	274.9	26.4	3349.5

TABLE IV
OVERALL INSTANCES OF MODES OF LOCOMOTION, ALONG WITH MINIMUM, MAXIMUM, AVERAGE AND TOTAL DURATION (SECONDS).

	#	Min Len	Max Len	Mean Len	Tot Time
Ambient	3426	0.2	6.3	1.0	3313.4
Objects	3709	0.2	64.5	2.0	7399.8

TABLE V
OVERALL INSTANCES OF HAND INTERACTIONS WITH THE OBJECTS AND INFRASTRUCTURE (SECONDS).

C. Data annotation

A dedicated open source tool was developed by the University of Passau to browse synchronously through the 3 video streams and the sensor data and mark the occurrence of relevant events (e.g. gestures, activities). The annotations are done on four 'tracks'. One track contains modes of locomotion (e.g. sitting, standing, walking). Two other tracks indicate the actions of the left and right hand (e.g. reach, grasp, release), with an attribute indicating to which object they apply (e.g. milk, switch, door). The fourth track indicates the high level activities (e.g. prepare sandwich). This level of annotation is sufficient for most applications, and allows to derive coarser annotations if needed. Annotation is executed by students. A 30-minutes video footage requires about 7-10 hours to be annotated.

V. EVALUATION OF THE DATASET ACQUISITION

A. Evaluation of activity instances

We annotated the activity occurrences from the video footage after the recording. Currently, 16 out of 60 ADL runs and 3 out of 12 drill runs are annotated.

In table IV we present statistics on the occurrences of modes of locomotion and table V 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 19 annotated runs. Extrapolating from this to the whole dataset, over 13000 interactions with objects and 14000 interactions with the environment may have been recorded. In table V we break down the hand interactions with objects in the 16 annotated ADL runs. In table VII we break down the right hand interactions during the 3 drill runs. As expected, activities occur roughly in multiples of 20 instances (subjects repeated 20 times the drill sequence).

By extrapolating from the currently labeled sessions, we estimate that the activity runs total 25 hours of recorded data. The cumulative length of all the annotations on all tracks represents approximately 57 hours of labels. These include posture/locomotion labels which are always present and hand interactions that often occur in overlapping fashion and including multiple objects, thus giving us a very label-intensive recording.

B. Evaluation of wireless data acquisition performance

A first assessment of the quality of the dataset is obtained through statistical measures of the amount of data that were

	reach	move	release	stir	sip	bite	cut	spread
Cup	24/103	26/171	19/91	0/4	4/67	0/0	5/11	8/31
Glass	12/64	17/103	11/57	0/11	1/39	0/0	0/1	0/0
Spoon	2/14	4/27	3/15	0/9	0/0	0/0	11/28	13/47
Sugar	16/19	22/19	19/13	0/0	0/0	0/0	15/17	21/22
Knife1	8/28	8/29	5/22	0/0	0/0	0/0	4/6	3/3
Knife2	10/29	13/37	9/24	0/0	0/0	0/0	3/22	0/7
Salami	24/40	37/43	22/36	0/0	0/0	0/0	2/26	0/0
Bottle	11/13	12/26	9/12	0/1	0/0	0/0	7/14	12/17
Plate	16/23	19/27	15/19	0/0	0/0	0/0	5/13	0/0
Cheese	26/25	38/26	23/20	0/0	0/0	0/0	42/16	0/6
Bread	62/68	105/125	53/56	0/0	0/0	25/36	13/47	15/72
Milk	24/36	21/33	15/24	0/0	0/0	0/2	4/9	5/20

TABLE VI

NUMBER OF INTERACTIONS OF THE LEFT/RIGHT HAND WITH OBJECTS.

	reach	open	close	lock	unlock
Fridge	84	43	43	0	0
Dishwasher	97	50	46	0	0
Drawer1 (top)	113	60	61	0	0
Drawer2 (middle)	108	59	62	0	0
Drawer3 (lower)	91	52	51	0	0
Door1	119	61	63	59	53
Door2	121	59	64	61	57
Switch	122	0	0	0	0
Table	61	0	0	0	0
Chair	3	0	0	0	1

TABLE VII

NUMBER OF RIGHT HAND INTERACTIONS WITH THE ENVIRONMENT
(DOORS, DRAWERS ETC) IN THE DRILL RUN.

lost during the acquisition process. During setup, we tuned the parameters of the wireless sensors. In particular for the Bluetooth motion sensors, we started from the highest sample rate (64 Hz) and transmission of all sensor channels (nodes can locally convert raw acceleration into calibrated values), and we progressively reduced the sample rate and eliminated information that could be recovered during post-processing (table VIII). Eventually, stage 3 settings correspond to the available bandwidth with DM1 ACL packets (max throughput 108.8kbps), not accounting for RFCOMM retransmissions. All 24 Bluetooth sensors streamed data simultaneously to 6 dongles connected to the recording computers.

Stage (runs)	Sample rate [Hz]	Data	Packet size [Byte]	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 VIII

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 these unfavorable conditions and the large amount of wireless devices, the overall packet loss was quite low after tuning. In particular, packet loss dropped to an average of 2.5% in stage 3 (table VIII), which is 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 6, where we can see that the sensors were streaming with little packet loss and only some
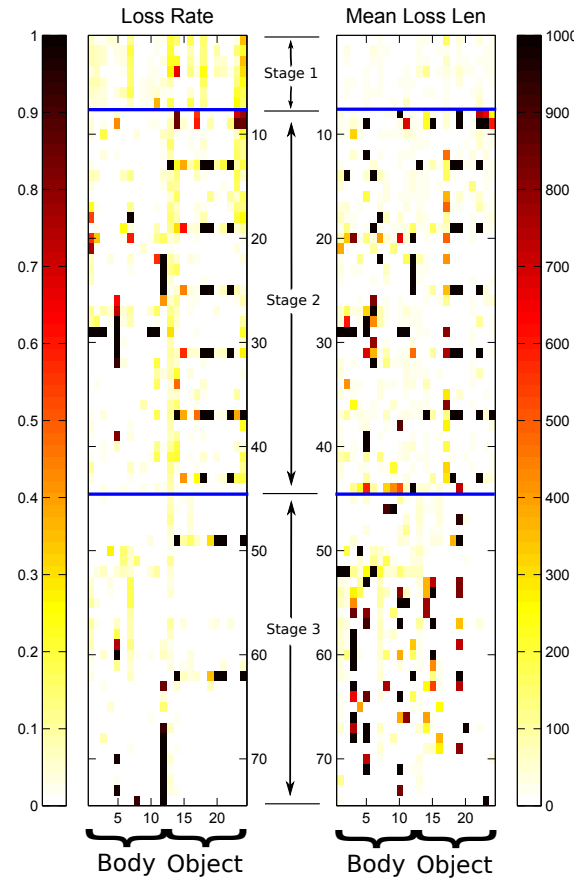

Fig. 6. Loss rate (left plot) and mean disconnection length (measured in samples). Disconnection length values have been saturated to 1000 to improve the plot clarity. In the scripted runs, some sensors have been switched off because not used, this can be seen from the regular black spots on the graphs.
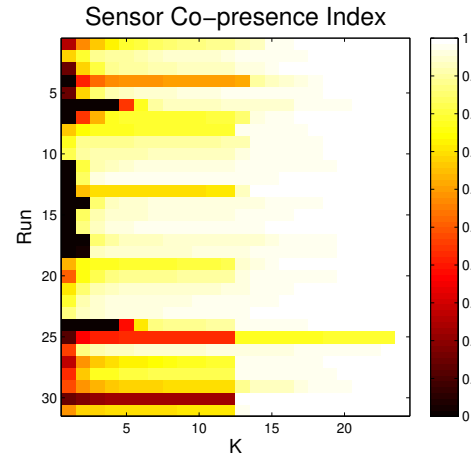

Fig. 7. Fraction of the data streams where at most K sensors are missing at the same time. The leftmost column for example indicates for how much time all sensors were present all together.

sporadic exceptions occurred. In most of the cases the average length of data losses was below 30 samples (right plot).

For a reference dataset it is important that more sensing modalities are present at the same time. In figure 7 (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 represents 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, nearly all the sensors streamed 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.

Further characterization may include higher-order statistics. Another characterization of this dataset for activity recognition may be the information content present in each sensor channel, thus allowing to quantify the contribution of each sensor to discriminate a set of activities.

VI. LESSON'S LEARNED AND BEST PRACTICES

Our experience with using `unison` and `subversion` for managing our dataset internally is satisfactory. It is important to ensure regular backups of the `unison` repository (done server-side with no network overhead) as multiple users are accessing the same repository. It is important to give clear instructions and procedures to the users to avoid inadvertent errors (e.g. file deletion - which did not happen so far).

The scenario described here differs substantially from typical homogeneous WSN and tends to require dedicated solutions, e.g. here the data synchronization was done during post processing. Eventually, protocols such as 6LoWPAN may offer a uniform end to end access to sensors.

We did not find existing tools for synchronized exploration and annotation of our dataset. Some were only partially applicable (e.g. limited to a single camera view). It has been worthwhile to invest time to develop custom tools in-house. The effort for labeling can now be reduced as the tools are well integrated and allow us to hire students this purpose.

During a long recording in a complex setup it is likely that faults occur. We experienced expected sporadic wireless data loss (e.g. due to body occlusion) and in a few cases battery failures. Finally, mechanical connectors may fail (occurred once). Continuous monitoring of the sensor systems during acquisition as well as sporadic checks (e.g. once per day/subject/run) can find these problems. Upon problem detection, one of the following actions can be taken: continue the recording; fix the problem and resume the recording; fix the problem and restart the recording. The choice is a trade off between: i) time the subject spends in a recording; ii) how often data is checked for validity; iii) the seriousness of the problem. In our scenario, subjects performed for 4 to 7 hours. More thorough checks would make this even longer and not feasible anymore. Less thorough checks however may lead to undetected corrupt recordings. We chose to perform rapid checks between the runs and more thorough checks after each subject. Rapid checks consisted of monitoring file sizes, which are roughly similar for identical sensors from run to run and subject to subject. More thorough checks (packet loss analysis) were performed after each subject and at the end of the day. Some simple problems that were identified as they occurred (e.g. a sensor that detaches from the clothing) were fixed on the fly and the recording resumed. Improvements in monitoring tools obviously reduce the time required for data checks, however the trade off indicated above remains.

In order to minimize data loss, we designed the data acquisition softwares and the recording architecture so that individual sensor failures did not have a catastrophic effect on the other ones. In particular, when a wireless sensor (typically the Bluetooth sensors) lost connection the system automatically attempted to reconnect to that sensor. This allowed us to minimize the amount of data loss, as disconnections were not uncommon, but usually the connection resumed after just a few seconds.

When integrating many wireless systems, it is difficult to predict the overall behavior even when individual systems are well characterized. For instance, the Sun SPOT system performed well in isolation. When including the Bluetooth sensors the performance of both systems deteriorated until we adjusted their parameters (packet size and sample rate). Thus, enough time must be reserved for integration tests. Ideally, they should be performed on the experiment site to be as realistic as possible, and e.g. determine the best placement for antennas.

The motion jacket was effective to easily deploy inertial sensors on the body. The remaining sensors were attached to the body individually, which took most of the setup time (tot. about 30mn). While arguably being far from comfortable, users nevertheless kept their freedom of movement despite the large number of installed sensors. They could execute fine motor activities (e.g. spreading cheese on the bread) and took many postures indicating the system did not hamper them much (e.g. kneeling down to reach objects in the fridge, stretching arms and body to reach a cup high on a shelf, sitting, lying). The motion jacket helped in this respect, and we avoided blocking limb joints with the additional attached sensors. A garment with slots for all the sensors should be systematically considered for recordings of this scale as it decreases set-up time, and improves wearability.

The logistical aspects should not be underestimated (e.g. renting a kitchen-equipped room for 11 days in a University). Here, the set-up of the ambient infrastructure took 3 days and integration tests and optimization of the wireless networks parameters took 2 additional days.

Sensing body motion using streaming sensors tends to stress the wireless infrastructure, leading to packet losses. In order to acquire a reference data set, local node storage should be considered, with the wireless links reserved for synchronization, system status monitoring, and post-hoc data download. This would reduce or eliminate data loss. On the other hand, our recording allowed us to collect information about the typical failure modes of our sensors when deployed in large numbers.

VII. CONCLUSION

We described the aquisition of a dataset of human activity from 72 sensors of 10 different modalities and grouped in 15 different wireless and wired sensor network systems (proprietary and custom) integrated in the environment, in objects, and on the body. Overall, 28 wireless sensors operating in the 2.4GHz band were deployed in close proximity. This

heterogeneous sensor network architecture highly multimodal and sensor rich and geared toward activity recognition is a specificity of our work.

Twelve subjects executed activities of daily living in this environment, yielding an average of 2 hours of effective data per subject, for a total of 25 hours of sensor data. We have annotated 19 runs out of 72. The overall length of the label track is 2.3 times the effective length of the runs, and we estimate that over 13000 and over 14000 interaction primitives respectively with objects and the environment were recorded. Thus, this dataset is highly rich in gesture instances and annotations. To the authors' knowledge, this dataset is the largest to date for multimodal activity recognition.

The complexity of the integration led us to address many issues which we documented here in the form of lessons learned and best practices. We presented the first results of the evaluation of wireless sensor data acquisition. After parameter tuning, packet losses were 2.5% for the 24 streaming Bluetooth wireless sensors and we achieved a high sensor co-presence index. Given the complexity of the wireless network deployment, this is very satisfactory for the purpose of activity recognition.

Ongoing work includes the annotation of the rest of the dataset and the technical evaluation of other sensor modalities. We will use this dataset to validate methods we have developed to date for activity recognition in opportunistic sensor setups.

This activity dataset is also well suited e.g. to investigate sensor fusion methods, to benchmark activity recognition methods or to do ontology-based reasoning. We invite members of the community to contact the OPPORTUNITY consortium to discuss an early access to this dataset.

REFERENCES

[1] L. Benini, E. Farella, and C. Guiducci, "Wireless sensor networks: Enabling technology for ambient intelligence," *Microelectron. J.*, vol. 37, no. 12, pp. 1639–1649, 2006.

[2] R. Verdone, *Wireless Sensor And Actuator Networks: Technologies, Analysis And Design*. Academic Press, 2008.

[3] M. Tentori and J. Favela, "Activity-aware computing for healthcare," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 51–57, 2008.

[4] T. Stiefmeier, D. Roggen, G. Ogris, P. Lukowicz, and G. Tröster, "Wearable activity tracking in car manufacturing," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 42–50, 2008.

[5] L. Holmquist et al, "Building intelligent environments with smart-its," *IEEE Computer Graphics and Applications*, vol. 24, no. 1, pp. 56–64, Jan./Feb. 2004. [Online]. Available: http://www.smart-its.org

[6] J. Beal and J. Bachrach, "Infrastructure for engineered emergence on sensor/actuator networks," *IEEE Intelligent Systems*, vol. 21, no. 2, pp. 10–19, 2006.

[7] F. Dressler, *Self-organization in sensor and actor networks*. Wiley, 2007.

[8] D. Roggen, K. Förster, A. Calatroni, T. Holleczek, Y. Fang, G. Tröster, P. Lukowicz, G. Pirkl, D. Bannach, K. Kunze, A. Ferscha, C. Holzmann, A. Riener, R. Chavarriaga, and J. del R. Millán, "Opportunity: Towards opportunistic activity and context recognition systems," in *Proc. 3rd IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications*, 2009.

[9] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and Systems Magazine*, vol. 6, no. 3, pp. 21–45, 2006.

[10] D. Bannach, O. Amft, and P. Lukowicz, "Rapid prototyping of activity recognition applications," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 22–31, 2008.

[11] C. Lombriser, M. Rossi, A. Breitenmoser, D. Roggen, and G. Tröster, "Recognizing context for pervasive applications with the Titan framework," Wearable Computing Laboratory, ETH Zurich, Tech. Rep. 624, 2009. [Online]. Available: http://e-collection.ethbib.ethz.ch/view/eth:624

[12] G. Fortino, A. Guerrieri, F. L. Bellifemine, and R. Giannantonio, "SPINE2: Developing BSN applications on heterogeneous sensor nodes," in *Proc. IEEE Symposium on Industrial Embedded Systems SIES2009*, 2009.

[13] S. Intille, K. Larson, E. Tapia, J. Beaudin, P. Kaushik, J. Nawyn, and R. Rockinson, "Using a live-in laboratory for ubiquitous computing research," in *Proc. Int. Conf. on Pervasive Computing*, 2006, pp. 349–365.

[14] T. van Kasteren, A. Noulas, G. Englebienne, and B. Kröse, "Accurate activity recognition in a home setting," in *Proceedings of the 10th international conference on Ubiquitous computing*. ACM Press, 2008, pp. 1–9.

[15] T. Huynh, M. Fritz, and B. Schiele, "Discovery of activity patterns using topic models," in *Proceedings of the 10th international conference on Ubiquitous computing*. ACM New York, NY, USA, 2008, pp. 10–19.

[16] K. Van Laerhoven, H. Gellersen, and Y. Malliaris, "Long-term activity monitoring with a wearable sensor node," in *BSN Workshop*, 2006.

[17] M. Tenorth, J. Bandouch, and M. 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) at ICCV*, 2009.

[18] K. Förster, D. Roggen, and G. Tröster, "Unsupervised classifier self-calibration through repeated context occurences: Is there robustness against sensor displacement to gain?" in *Proc. 13th IEEE Int. Symposium on Wearable Computers (ISWC)*, 2009, pp. 77–84.

[19] K. Kunze and P. Lukowicz, "Dealing with sensor displacement in motion-based onbody activity recognition systems," *Proc. 10th Int. Conf. on Ubiquitous computing*, Sep 2008.

[20] K. Förster, P. Brem, D. Roggen, and G. Tröster, "Evolving discriminative features robust to sensor displacement for activity recognition in body area sensor networks," in *Proc. 5th Int. Conf. on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP 2009)*. IEEE Press, 2009.

[21] K. Kunze and P. Lukowicz, "Using acceleration signatures from everyday activities for on-body device location," *Wearable Computers, 2007 11th IEEE International Symposium on*, pp. 115 – 116, Sep 2007.

[22] ——, "Symbolic object localization through active sampling of acceleration and sound signatures," *Ubicomp*, 2007.

[23] A. Calatroni, C. Villalonga, D. Roggen, and G. Tröster, "Context cells: Towards lifelong learning in activity recognition systems," in *Proceedings of the 4th European Conference on Smart Sensing and Context (EuroSSC)*. Springer, 2009, pp. 121–134.

[24] K. Förster, A. Biasiucci, R. Chavarriaga, J. del R. Millàn, D. Roggen, and G. Tröster, "On the use of brain decoded signals for online user adaptive gesture recognition systems," in *Proc. 8th Int. Conf. on Pervasive Computing*, 2010.

[25] D. Roggen, M. Bächlin, J. Schumm, T. Holleczek, C. Lombriser, L. Widmer, D. Majoe, J. Gutknecht, and G. Tröster, "An educational and research kit for activity and context recognition from on-body sensors," in *Proc. Int. Conf. on Body Sensor Networks (BSN)*, 2010.

[26] G. Pirkl, K. Stockinger, K. Kunze, and P. Lukowicz, "Adapting magnetic resonant coupling based relative positioning technology for wearable activitiy recogniton," in *12th Int. Symposium on Wearable Computers (ISWC)*, 2008, pp. 47–54.

[27] P. Zappi, C. Lombriser, E. Farella, D. Roggen, L. Benini, and G. Tröster, "Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection," in *5th European Conf. on Wireless Sensor Networks (EWSN 2008)*, R. Verdone, Ed. Springer, 2008, pp. 17–33.

[28] T. Westeyn, P. Presti, J. Johnson, and T. Starner, "A naïve technique for correcting time-series data for recognition applications," in *Proc. 13th Int. Symposium on Wearable Computers*, 2009, pp. 159–160.