Evaluation of localization and activity recognition systems for ambient assisted living: The experience of the 2012 EvAAL competition

Juan Antonio Álvarez-García [a,*], Paolo Barsocchi [b], Stefano Chessa [b,d] and Dario Salvi [c]

[a] *Computer Languages and Systems Department, University of Seville, Spain*
[b] *ISTI-CNR, Pisa Research Area, Pisa, Italy*
[c] *Life Supporting Technologies, Technical University of Madrid, Spain*
[d] *Computer Science Department, University of Pisa, Italy*

Abstract. EvAAL is an annual international competition that addresses the "grand" challenge of evaluation and comparison of Ambient Assisted Living (AAL) systems and platforms, with the final goal to assess the autonomy, independent living and quality of life that AAL systems may grant to their end users. The 2012 Edition was focused on two pillars of AAL: Indoor localization and activity recognition. Results from both competitions suggest that there is still space for other editions not only to improve accuracy of such systems, but also their user acceptance and interoperability. This paper describes the organization and results of the 2012 edition.

Keywords: Indoor location, activity recognition, AAL

1. Introduction

Ambient Assisted Living (AAL) [1], a research program of the EU that addresses ICT technologies for the independent living of elders and disabled [2], has attracted the attention of a large community of researchers, industries and users [3]. In particular, one of the main interest for researchers is in the development of platforms for AAL, that should be able to integrate a large number of heterogeneous devices, functionalities and services, along with features of dependability, human/machine interaction etc. In other words, AAL is a highly interdisciplinary field that involves many different areas of ICT.

Despite the effort of research in AAL, the evaluation and comparison of AAL solutions and platforms is far from being a reality [4]. Complexity of such solutions and the mix between quantitative and qualitative aspects in the evaluation that includes performance, availability, user acceptance or interoperability caused this failure.

It is in this framework that the EvAAL competition (standing for Evaluating AAL Systems through Competitive Benchmarking) was born [5]. EvAAL is an annual international competition promoted by the AALOA association [6] that addresses the "grand" challenge of evaluation and comparison of AAL systems and platforms, with the final goal to assess the autonomy, independent living and quality of life that AAL systems platforms may grant to their end users. To this purpose EvAAL supports the creation of a community of stakeholders in AAL (including researchers, industries, service providers and end users) that can openly discuss these issues and test their ideas and methodologies in a public competition. The techni-

*Corresponding author. E-mail: jaalvarez@us.es.

cal objectives of the competition are not limited to the comparison and evaluation of different AAL solutions, but they also include the development and experimentation of benchmarks and evaluation methods and the identification of relevant open problems in AAL to be addressed in the forthcoming editions. An important collateral result of EvAAL is the creation of an open knowledge base containing evaluation data, benchmarks and software tools developed in the various editions of EvAAL.

Recognizing that the comparison and evaluation of full AAL platforms is not feasible with the state of the art, EvAAL adopts a bottom up approach: it initially develops the methodologies for evaluation of simple services and components of AAL systems; once these methodologies are developed and mature, it focuses on the evaluation of aggregated components and services, and finally entire platforms. Currently EvAAL is in the first phase (evaluation of simple components). The first edition in 2011 focused on the indoor localization problem, the 2012 edition included also a track on activity recognition, and in the next edition it is planned to extend EvAAL with additional tracks in future editions. The objective is to organize competitions on more complex AAL services on a time horizon of five years. This paper describes the organization and results of the second (2012) EvAAL edition.

The rest of the paper is organized as follows: Section 2 gives a general overview of EvAAL 2012 and Section 3 reviews some related competitions. The technical details of the tracks on indoor localization and on activity recognition are presented in Sections 4 and 5, respectively. Finally, Section 5 draws the conclusions.

2. Overview of EvAAL 2012

The 2012 EvAAL competition followed the same organization of the 2011 edition. In particular, the first step was the publication of a call for ideas: a public call aimed at stimulating an open discussion about the topics of interest and at defining the focus of the forthcoming edition. Following the feedbacks to the call from the community, and considering that EvAAL is still in its first phase in which simple components are evaluated, it was decided to organize two tracks based on the topics of indoor localization and of activity recognition.

The first track took place on the 2–6 July, 2012 at the Living Lab of the Polytechnic University of Madrid

[7], and the second took place on the 9–13 July, 2012 at the CIAmI Living Lab [8] in Valencia in Spain. The winners had been announced during the concluding EvAAL workshop, that was held on the 24th September 2012 in conjunction with the AAL Forum [3] in Eindhoven.

The track on indoor localization and tracking for AAL was chosen because it is a key component of many AAL services, that has also attained an increasing attention by the research community motivated by the need for location-based services and applications. This track was already present in the 2011 edition, but it was improved based on the experience of the previous year. The track on activity recognition instead, covers a less mature problem that, nevertheless, represents an important enabling feature in many AAL services. In fact, real-time monitoring of human activities represents a useful tool for many purposes and future applications such as lifelog, healthcare or entertainment. On the other hand, the automatic and unobtrusive identification of user's activities is one of the challenging goals of context-aware computing.

2.1. Contribution

The EvAAL event is a live competition, meaning that the competing artifacts are evaluated in a real home environment in almost-real settings. Competitors are requested to install and run their systems during a set of benchmarks, within a time frame of three hours. An Evaluation Committee (EC) is present during the competition to control all the operations and to ensure a fair evaluation of each artifact. The time slot assigned to each competitor is divided in three parts:

- In the first part, a competing team deploys and configures its artifact in the living lab. This part lasts no more than 60 minutes and its duration is measured in order to produce the score for installation complexity criteria (see Section 4.4.1).
- In the second part, the benchmarks are applied. The benchmarks are the same for each competing team and are not disclosed in advance to the teams, so that no competing teams is favoured. During this phase the competing team can, if necessary, perform only short reconfigurations of its systems. In any case, this part lasts no longer than one hour.
- In the last part, the competing team removes its system from the living lab, and the evaluation committee restores the initial condition of the living lab for the next competing team.

3. Related competitions

Competitions in computer science have a long-established tradition. The design of EvAAL has been inspired by other successful competitions. Some past and current competitions have been analyzed in order to identify successful practices, specifically: the Trading Agent Competition [9], the DARPA Grand Challenge [10], the International Collegiate Programming Contest [11], the International Olympiad in Informatics [12], the Google Code Jam[1] and the CONNECT Code-a-Thon Challenge[2]. Below we describe some of these.

HARL [13] The Human activities recognition and localization competition (HARL) is a competition launched within the International Conference on Pattern Recognition (ICPR) in 2012. The goal of HARL is to recognize complex human activities involving several people at the same time interacting among themselves and with artifacts. Competitors have to analyze a set of data collected by two cameras, a moving Kinect camera mounted on a mobile robot and a high-resolution, color, fixed camera, and classify 10 activities based on these data. Competitors have to provide an executable which takes the video as input and provides an XML file with the results as output. The ranking of the solutions is conducted offline by executing metrics on the XML file. Both the number of correctly recognized activities as well as their location are considered for the definition of the final ranking.

OPPORTUNITY [14–16] Is an extremely sensor-rich and activity-rich common dataset against which all participants benchmark their proposed activity recognition software. The dataset includes 72 body-worn, ambient, and object sensors, a very high number of activity instances (more than 2500 instances of gestures) labeled at various levels of abstractions, executed by multiple persons. The dataset is available on the UCI ML repository A new challenge is in preparation for 2013 at the International Symposium on Wearable Computers.

HASC [17,18] The main purpose of HASC is to collect a large scale human activity corpus of 6 activities: stay, walk, jogging, skip, stair-up, stair-down. HASC mainly uses inertia sensors such as accelerometer, gyroscopes, and geomagnetic sensors, and allows GPS or other sensor available in the market although major types of sensors are iPhone/iPod Touch, and WAA-series (ATR). HASC2011corpus is composed by 4898 sensor data files from 116 subjects and it is publicly available to the research community.

HASC sets several levels of contribution. Participants can contribute with new sensor data for these activities, in exchange of having access to the entire database of data, which is the first level of contribution. This leads to an organically "crowd-sourced" growth of the dataset. The second level of contribution consists in a competition on activity recognition software, where the competitors compete with systems aimed at recognizing the HASC dataset.

BSN contest [19,20] The main purpose of this contest is to share BSN medical data among researchers and to compare activity recognition algorithms on the same data set. Five sensors were used (ankle, tight, chest, belt and wrist worn as wristwatch), each of them has a 3-axis gyroscope and a 3-axis accelerometer. The contest contained the following tasks:

1. Signal segmentation and movement recognition for the following transitions: sit to stand, stand to sit, sit to lie, lie to sit, turn counter-clockwise 90 degrees, turn clockwise 90 degrees, pick a book up from the floor, place a book on the shelf, step forward (1 step).
2. Stride time assessment including treadmill and free walking.
3. Sit-to-stand recognition.

4. Indoor localization track

4.1. Motivation

Localization is a key component of many AAL systems, since the user position can be used for detecting user's activities, activating devices, opening doors, etc. While in outdoor scenarios Global Positioning System (GPS) constitute a reliable and easily available technology, in indoor scenarios GPS is largely unavailable.

For this reason, several systems have been proposed for indoor localization. These systems can be classified based on the signal types (infrared, ultrasound, ultra-
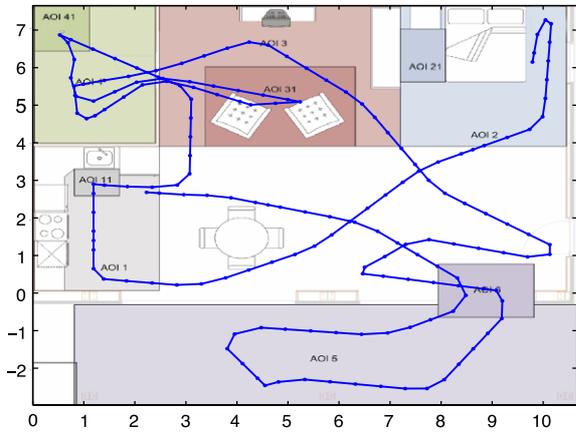
---

[1] http://code.google.com/codejam
[2] http://hit.fiu.edu/challenge.htm

wideband, and radio frequency), signal metrics (AOA – angle of arrival, TOA – time of arrival, TDOA – time difference of arrival, and RSS – received signal strength), and the metric processing methods (triangulation and scene profiling) [21]. Each solution has advantages and shortcomings, which, in most cases, can be summarized in a trade-off between precision and installation complexity (and thus costs).

In practice, although indoor localization has been a research topic for several decades, there is still not a de-facto standard. Moreover, localization in AAL applications has specific requirements due to the fact that AAL systems must be deployed in homes. In particular, localization system for AAL should be well hidden, easy to install and configure, and reliable. For these reasons EvAAL includes a track on indoor localization.

The call for competition for this track, launched on January 2012 on the EvAAL website[3] and now closed, opens to all localization technologies that can be deployed in homes. For this reason the benchmarks defined in this track are built making no assumption on the technologies used in the competing artefacts. The organization of this track also includes the preparation of a reference localization system and of the evaluation tools, that are aimed at collecting in real-time the localization data from the competing systems, and at running a set of metrics on the competing systems output to produce the final scores.

4.2. Benchmarks description

The track on indoor localization is a real-time competition, meaning that a real human actor has to be localized in a real home scenario. The competition was hosted by the Smart Home Living Lab, at the Technical University of Madrid in Spain.

During the benchmark phase, an actor (a member of the evaluation committee) wears the equipment the competitors requires to carry (if any) and moves along a set of predefined paths (the chosen paths are represented in Fig. 1). While moving, the localization data produced by the localization system of the competitor are collected in real time by the data collection tool.

The competing systems have also access to the domotic equipment of the Living Lab, which includes configurable switches, lights, movement sensors, as

Fig. 1. The three different paths: path 1 (blue line), path 2 (red line), and path 3 (green line). (Color figure online)

well as electronic kitchen appliances. Hence the localization algorithms can exploit the information produced by these devices as consequence of the movements and actions of the actor. For this reason the benchmarks also include actions such as turning on/off lights or opening doors that are detected by the domotic equipment of the Living Lab and provided to the localization systems in form of "contextual" events.

The benchmarks are divided into three scenarios.

– Scenario 1. In this scenario the localization systems have to locate a person inside an Area of Interest (AoI). AoIs represent areas that can have a specific meaning in an AAL application. Examples of AoI can be specific rooms (kitchen, bedroom, etc.) or areas where appliances are located (close to the fridge, on the bed etc.). Each system is requested to identify 6 big AoIs (representing rooms) and 4 small AoIs (representing points of interest for the user). The actor moves along predefined paths and stops inside each AoI for at least 5 seconds. The AoI used in the benchmarks are shown in Fig. 2.

– Scenario 2. In this scenario, the actor has to be located while moving in the living lab along predefined paths. The expected output of the localization systems is the stream of his actual positions (in bi-dimensional coordinates) and the respective timestamps. During this phase only the person to be localized is inside the Living Lab. Each localization system is requested to produce localization data with a frequency of 1 sample every half a second. This benchmark uses two paths: one 54 steps long (path 1 in Fig. 1), and one 94

Fig. 2. The Areas of Interest deployed in the Smart House Living Lab and the path followed by the actor.
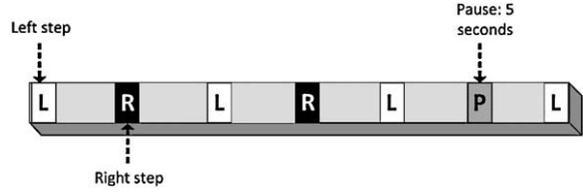

Fig. 3. The reference localization system: the black marks are related to the right foot while the white ones are related to the left foot. Marks denoted with "P" denote a stop of the actor for a given time.
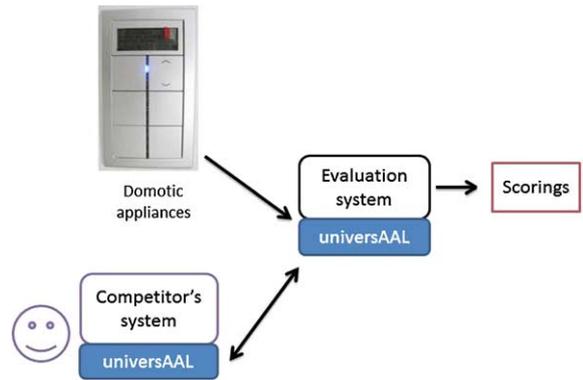

Fig. 4. Communication among competitor's node, the evaluation system and the Living Lab.

steps long (path 2 in Fig. 1). Each path includes 3 waiting points, where the actor has to stay still in the same position for 5 seconds.

– Scenario 3. This scenario is similar to Scenario 2, with the difference that another actor (a disturber) moves in the living lab together with the primary actor. In this scenario only the primary actor has to be localized as in the previous scenario. The disturbing actor follows different, predefined paths, also activating domotic equipment, but at least 2 meters away from the actor. In this scenario the paths followed by the actor are path 3 (80 steps long) and path 2, while the disturber paths are path 2 and path 1, respectively. These paths are shown in Fig. 1.

4.3. The reference localization system

The reference localization system is used to compare the localization data generated by the competitors with the ground truth. The reference consists in a set of pre-defined paths the actor has to follow with a predefined speed. As shown in Fig. 3, the Living Lab's floor is covered with marks (with different colors to distinguish the right and left foot) that indicate each single step the actor has to follow. In order to facilitate the installation and removal of the paths the marks are put on a wooden bar. The actor is synchronized by a digital metronome that indicates the right cadence (one beep one step), guaranteeing that the actor repeats the same paths at almost the same speed for every competitor.

4.4. The evaluation system

During the benchmarks, each participant produces a large data-sets that has to be compared against the reference system in order to evaluate the final score. The synchronization aspects are critical for computing the final scores, since these are based on real time information like availability or jitter and should be correlated with the current position in the reference system. To this purpose, this track uses the *evaluation tool* (whose structure is shown in Fig. 4), which is responsible for:

– synchronizing the actor's movements by generating the sounds at each step and signaling pauses (like a metronome);
– generating the ground truth dataset for each scenario, using the description of each path and the AoIs;
– receiving the data from the competitors and storing it;
– connecting to the domotic appliances of the lab (based on the KNX protocol) and routing contextual events to the competitors;
– running the evaluation metrics and computing scores for each competitor;

– visualizing the ground truth current position and the one sent by the competitor (used for debugging and entertainment during the benchmarks).

In order to enable interoperability between the evaluation tool and the competing systems, before the competition the competitors are provided with an "integration package" that allows them to integrate their software with the EvAAL one. The package contain the necessary libraries to send localization data and receive contextual events, and it works both as a Java library or a stand alone application with a local socket interface.

The evaluation tool is developed on top of the UniversAAL [22] platform. universAAL is based on a semantic framework that helps integrating heterogeneous distributed systems, it resolves problems like node discovering and inter-node communication, as well as more advanced features including multi-modal user interaction and the installation and deployment of AAL applications. In this edition of the competition, only basic communication features of UniversAAL were used, although it is planned that further competitions will employ broader use of the platform.

Although the evaluation tool and the competing systems can send and receive data in real-time, an extensive log is generated on both machines in a human readable textual format. This feature allows the evaluation system to compute the metrics offline, in case of communication problems. Specifically the logs contain, for each participant and benchmark scenario, the following information:

– Benchmark name and timestamps of the start and end;
– Scenarios' descriptions, including the coordinates of every step of every path and the AoIs' coordinates;
– Coordinates of the localized actor (x and y axis) measured in meters;
– Instant of time when the event has been generated at the competitor's machine;
– Instant of time declared by the competitor (competitors could send corrected timestamps if errors or delays were predictable);
– Area of interest detected by the competitor;
– Ground truth data, generated in real-time thanks to the metronome information and the scenarios' descriptions;
– A graph showing the ground truth points and the estimated points, every second, plotted on the map of the lab;

– The values of each computed metric;
– Some extra debugging information.

4.4.1. Evaluation criteria

In order to evaluate the competing localization systems, the localization track uses a set of criteria weighted according to their relevance and importance for AAL applications. The localization track uses five criteria:

– *Accuracy*. It expresses the degree with which the competing system is able to correctly localize the actor. Accuracy is computed as the error distance between each localization sample sent by the competitor and the reference position. Accuracy is evaluated in two different ways:

  ∗ In scenario 1 accuracy is measured as the fraction of time in which the localization system provides the correct information about presence or not in a given AoI. The number of correctly guessed AoIs is averaged on the number of guessable AoIs.
  ∗ In scenarios 2 and 3 the euclidean distance between the coordinates sent by the competitor and the reference position is computed at every sample, then the 75th percentile of the errors is computed.

– *Availability*. Represents the fraction of time the localization system is active and responsive. The availability is measured as the ratio between the number of produced localization data and the number of expected data. Each localization system is expected to provide one sample every half a second, hence the number of expected samplings is given by the double of the test duration in seconds.

– *Installation complexity*. It measures the effort required to install the AAL localization system in a home. It is measured as a function of the person-minutes of work needed to complete the installation. The time is measured in minutes from the time when the competitors enter the living lab to the time when they declare the installation complete. The time of the main installer is accounted entirely, while the time spend by other installers are accounted by half of their time.

– *User acceptance*. It expresses how much the localization system is invasive in the user's daily life and thereby the impact perceived by the user. This parameter is estimated with a simple questionnaire that considers aspects of usability like

the presence and invasiveness of the tags, the visibility of the installation within the environment and the complexity of maintenance procedures.

– *Interoperability*. Measures how much the system is easy to integrate with other systems. This parameter is fundamental in AAL scenarios, as localization can be exploited by other applications to offer advanced services. Interoperability is measured with a questionnaire that takes into account aspects like the availability of APIs and documentation, the licensing scheme, the presence of testing tools and the portability among different operating systems.

All these metrics are normalized to a common 0 to 10 scale and mixed with a weighted average. The applied weights are: Installation complexity: 0.15, User acceptance: 0.25, Accuracy: 0.25, Interoperability: 0.15, Availability: 0.2. Accuracy and Availability are averaged over the three scenarios.

4.5. Contestants and technology

Seven teams were accepted to the indoor localization competition, namely CAR (from the Centre for Automation and Robotics, Spain) [23], LOCOSmotion (from the University of Duisburg-Essen, Germany) [24], OwlPS (from the Institute Femto-st, France) [25], CPS Group @ Utah (from the University of Utah, USA) [26], TAIS (from the University of Sevilla, Spain) [27], iLocPlus (from Stuttgart University of Applied Sciences and iHomeLab at Lucerne University of Applied Sciences) [28], and Smart-Condo (from the University of Alberta, Canada) [29]. The description of these systems is as follows:

CAR, Spain This system is based on the fusion of two complementary technologies: i) Inertial integration and ii) RFID trilateration. The Inertial solution uses an IMU (Inertial Measurement Unit) mounted on the foot of the person. The IMU approach estimates the user's trajectory shape; however, being a dead-reckoning method, it requires an initialization in position and orientation to provide absolute positioning. The RFID-based localization system provides the absolute position exploiting the Received Signal Strength (RSS) from several tags installed in the environment. The output of both IMU- and RFID-based methods are integrated into a fused position estimation.

LOCOSmotion, Germany It is an acceleration-assisted WLAN-based tracking system. Similar to many other systems, the basis of LOCOSmotion is fingerprinting technique. In order to achieve high update rates and to capture movements, this system augments the fingerprinting information with acceleration measurements.

OwlPS, France This system is WLAN-based localization system, that exploits the RSS to infer the user's position. OwlPS can be used as both fingerprinting-based system or trilateration-based system.

CPS group @ Utah, USA It is a device-free localization and tracking system, where people to be located do not carry any device. It is based on a wireless sensor network which uses a tomographic approach to localize the users. A static deployed wireless network measures RSS on its links and locates people based on the variations caused by the movements of people.

TAIS, Spain This system is based on RSS fingerprinting technique. The Manhattan distance metric is used to evaluate the user's position.

iLocPlus, Germany/Switzerland iLocPlus is an ultrasound ranging based indoor localization system. The user carries a badge including an ultrasound transmitter. This badge can be localized by means of at least three reference nodes (which position is a priori known) deployed in the environment.

Smart-Condo, Canada The Smart-Condo localization component relies on knowledge of both the coordinates and the mounting angle of where the motion sensors have been placed to construct a map of the space (segmented in a number of polygons). Given a pattern of sensor readings the localization system can be configured to use two methods for estimating the user's position. The first is a center-of-mass calculation of the motion sensors that are simultaneously triggered, the second is based on a probability distribution of the user's position in each of the polygons, considering the recent history of the estimated position.

4.6. Results

Table 1 summarizes the scores of the different competitors. Tables 1, 2, and 3 show the detailed score obtained by each team during the competition.

Table 1

The final scores and the jury decision scores (user acceptance and integrability, accuracy, availability and installation complexity)

Competitor	Accuracy	Availability	Installation complexity	User acceptance	Interoperability	Final score
CAR [23]	7.57	8.21	10	6.56	6.81	7.70
CPS Group @ Utah [26]	6.98	10	3.44	8.1	7.78	7.45
OwlPS[25]	0.78	10	9.71	6.39	6.92	6.29
Smart-Condo [29]	2.81	9.06	1	6.85	6.90	5.41
LOCOSmotion [24]	0.64	9.96	1.77	7.23	6.74	5.23
iLocPlus [28]	3.64	9.84	0	4.90	5.04	4.86
TAIS [27]	0.67	10	0	5.11	5.15	4.22

Table 2

The accuracy score

Competitor	Scenario 1 – AOIs	Scenario 2 – path 1	Scenario 2 – path 2	Scenario 3 – path 3	Scenario 3 – path 2
CAR [23]	6.27	8.96	8.02	5.89	8.68
CPS Group @ Utah [26]	8.78	8.95	9.39	7.79	0
iLocPlus [28]	3.45	3.56	4.86	5.02	1.28
Smart-Condo [29]	5.66	3.24	2.66	1.16	1.31
OwlPS[25]	3.11	0	0	0	0.82
TAIS [27]	3.35	0	0	0	0
LOCOSmotion [24]	3.21	0	0	0	0

Table 3

The availability score

Competitor	Scenario 1 – AOIs	Scenario 2 – path 1	Scenario 2 – path 2	Scenario 3 – path 3	Scenario 3 – path 2
CPS Group @ Utah [26]	10	10	10	10	10
OwlPS[25]	10	10	10	10	10
TAIS [27]	10	10	10	10	10
LOCOSmotion [24]	9.96	10	10	9.84	10
iLocPlus [28]	9.26	10	10	10	9.94
Smart-Condo [29]	9.10	8.43	9.63	9.10	9.06
CAR [23]	8.30	8.06	7.39	8.14	8.94

In particular, the CAR system reached the best overall score, since it received the best score for accuracy (Table 1) and installation complexity. As highlighted in Table 1, the CPS Group@Utah took the second place since it was the best system concerning availability (see Table 3, together with OwlPS and TAIS) and interoperability. In particular, this system was very accurate in all the tests except for the second path of the third scenario (Table 2). In this test the system performed poorly due to the disturbing actor, since it has no mean to discriminate among different actors. Figures 5 and 6 show a qualitative representation of the accuracy of the CAR and the CPS Group@Utah systems. As shown in this figures the CPS Group@Utah system is the most accurate, since it reaches a very high precision with respect to the other systems. Being a device-free localization, its main strength is the user acceptance but since it requires many radio devices to locate the user, installation complexity is its main drawback. At the third place the OwlPS system reached an high score in installation complexity, availability, and interoperability but it was not particularly accurate. Figure 7 shows the Cumulative Distribution Function (CDF) of error for the first three teams namely CAR, CPS Group@Utah, and iLocPlus (Table 2). The CDF is evaluated for the two best performing path with and without the disturber. The teams that use a mobile terminal are not affected by the presence of a disturber that moves inside the living lab during the competition. As a matter of fact, the CAR and iLocPlus teams obtained a similar performance in the two scenarios, while the CPS Group@Utah is the most accurate sys-
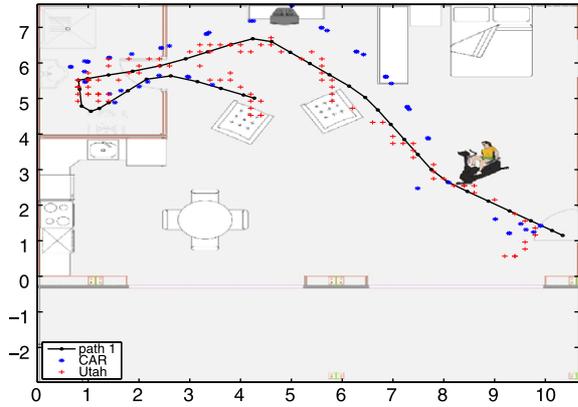
Fig. 5. The results of both CAR and CPS Group@Utah systems in the second scenario and first path.
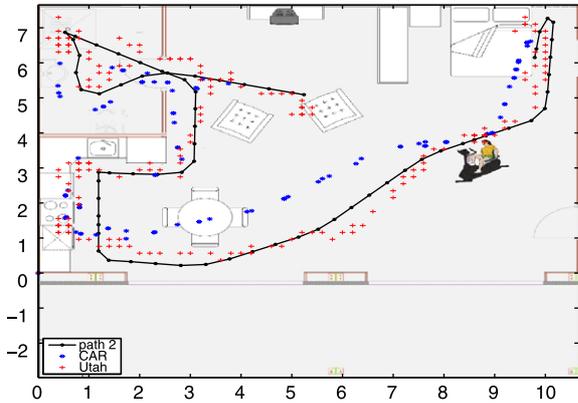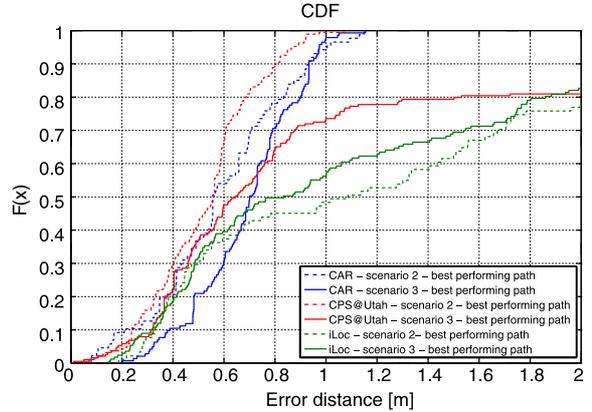

Fig. 7. The CDF of the distance error for the first three competitors with (solid line) and without the disturber (dotted line).

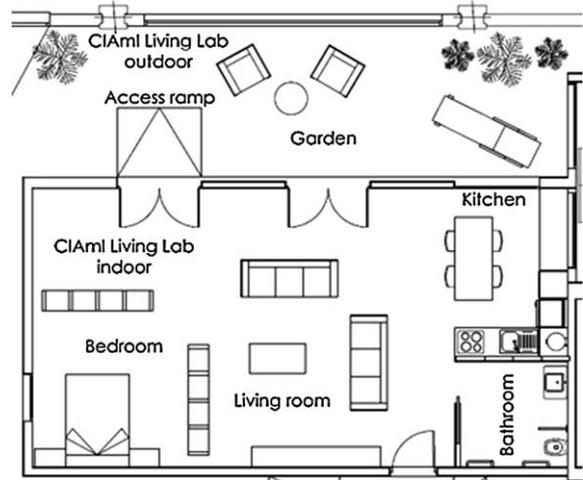Fig. 6. The results of both CAR and CPS Group@Utah systems in the second scenario and second path.

Fig. 8. Map from CIAMI Living Lab.

tem without a disturber, but its performance decrease significantly in Scenario 3 with the disturbing actor.

5. Activity recognition track

5.1. Motivation

The main objective of this track is to implement an activity recognition system (ARS) that recognizes the following activities: lie, sit, stand, walk, bend, fall and cycle (using a stationary bike).

In this track there is no limitation to the number of devices that can be used and competing solutions can be based on a variety of sensors and technologies, including: accelerometers, gyroscopes, magnetometers, pressure sensors, microphones, sensor networks, mobile phones, cameras, etc.. Other technolo-gies or combinations of them are also considered acceptable provided they are compatible with the constraints of the hosting Living Lab (CIAMI). Figure 8 shows the CIAMI map.

Although there already exist many competitions on activity recognition (as discussed in Section 3), the main difference with the preceding ones is that this track addresses both software and hardware so that whole state of the art of activity recognition can be examined.

5.2. Benchmark description

As described in Section 4.2 competitors are invited to install and run their ARS during a time slot, in this case for two hours and a half, divided in three subslots: Installation, Benchmark and Removal phases.

During the second phase, the ARS are evaluated. An actor (an evaluation committee member), performes a predefined physical activity trip across the smart home. Audio signals, explained in Section 5.3, synchronize the actor movements in each performance (twice per team) in order to get the same ground truth for all the participants. The path followed by the actor and the activities are the same for each performance, and they were not disclosed to competitors before the application of the benchmarks. Similarly, the position of the stationary bike and the place of the fall are not revealed either.

Once the two performances are executed, the one with better overall mark is used to compare each team.

A critical issue (that was communicated in advance to the competitors) is the age of the actor, that is required in order to train and prepare their algorithms properly. The actor was also trained to repeat the activities in the benchmarks always in the same way (following the mp3 file explained later). The fall is also critical, because many different kind of falls are possible [30]. For this reason we published in advance a video[4] of the fall that had been executed by the actor during the experiment.

5.3. Reference and evaluation system

The reference ARS is used to obtain the ground truth data. In order to get approximately the same ground truth for all the contestants, audio signals were used to synchronize the actor movements. An mp3 file indicates the next activity that the actor must perform and a countdown ("three, two, one, now") to perform it. When the actor hears the word "now", he begins the transition to the next activity. In some cases, such as BSN Contest, researchers identify transitions to recognize the next activity. In our case with 7 activities, the number of possible transitions (some of them not very probable) is 42. Since the number of transitions is high and it is not trivial to evaluate them, we decided not to evaluate the transitions but only the activities.

To retrieve competitor and ground truth data, a local server accepts sockets with the activity code and the time when it is identified. A competitor system only need to send this information to the server. A local NTP server is also available to synchronize the time.
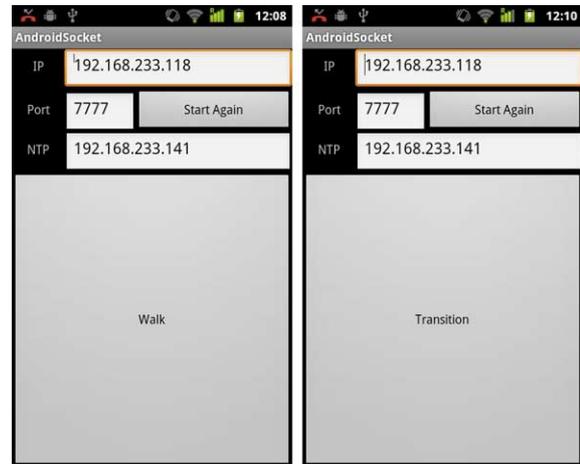
Fig. 9. Android application for label ground truth data.

To obtain ground truth data, an evaluator uses the Android application shown in Fig. 9 to mark the activities (activity code) or transitions (-1 code) and to send this information to the socket server.

To this purpose, the evaluator follows the actor to see exactly when he starts the activities or transitions. For instance, if the actor is standing still and hears "Cycling 3, 2, 1, now!" he begin to hold the handlebars and move his leg to go up the bike. When the word "now" is spelled, the evaluator pushes the button in the application to identify the end of previous activity and the beginning of a transition. The evaluator later on pushes again the button to mark the end of the activity.

5.3.1. Evaluation criteria

Almost the same criteria explained in Section 4.4.1 are used to evaluate the ARS with the same weights. An exception is Availability, which is replaced with Recognition delay (weight 0.2) that measures the elapsed time between the instant in which the user begins an activity and the time in which the system recognizes it. The accuracy criteria evaluates the recognized activity instances (500 ms) using F-measure $\frac{2*precision*recall}{precision+recall}$ to compute it.

5.4. Contestants and technologies

After peer review, five teams were accepted but one of them withdrew due to financial cutbacks of its institution. Hence only four competitors participated in the challenge, CUJ (from the University of Chiba, Japan) [31], CMU (from Carnegie Mellon and Utah Universities, USA) [32], DCU (from Dublin City Univer-
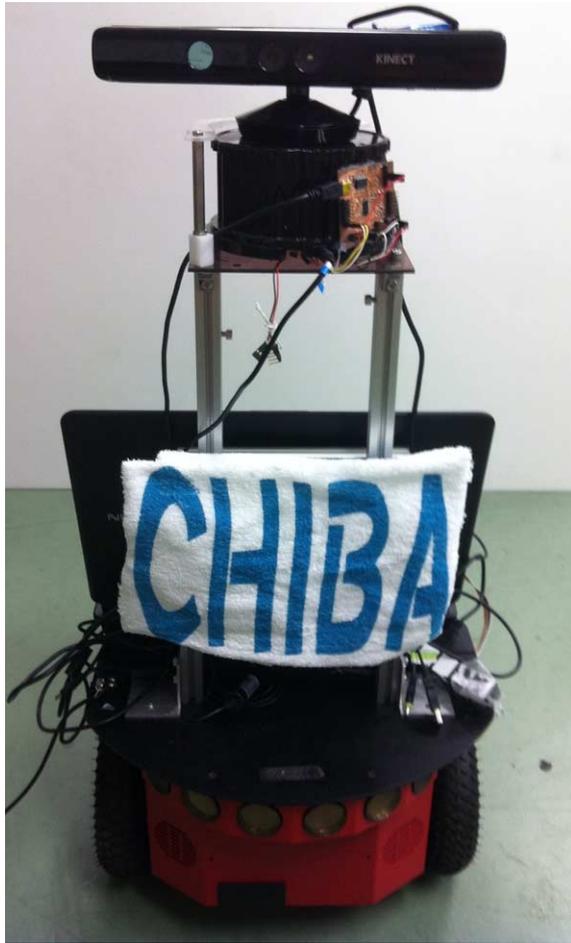
Fig. 10. Chiba team robot.

Fig. 11. Devices used by CMU and Utah Universities. Top image shows the chest strap and android mobile wore by our actor. Down at left the case where the radio beacons are stored can be seen. Down at right a detail where a radio was placed in a door.

sity, Ireland) [33] and USS (from University of Seville, Spain) [34]. The technology used by every team is described here:

CUJ uses a robot (initially a Pioneer 3-AT[5] but due to difficulties in transport from Japan, a Roomba[6] was used instead) with two kinects[7]. The first Kinect is used to avoid obstacles when following the actor and the other to recognize their activities. Figure 10 shows the evolved robot after the competition with only one Kinect and the Pioneer as platform.

CMU proposes a solution composed by three subsystems: A chest wearable elastic strap[8] capable of measuring several physiological signals, an Android mobile phone and a system for indoor localization based on Radio tomographic imaging[9]. Figure 11 shows the devices used by this team.

DCU uses a SenseCam[10] hanging from the actor's neck to evaluate off-line the activities.

USS uses an android mobile phone placed on the right hip. The user activities are recognized by means of the accelerometer embedded in the mobile phone.

5.5. Results

Table 4 shows the results for each team and criteria for the best of both performances. The winner (USS team) obtained acceptable results in accuracy (it was below that of the CMU team), but its simplicity (although it uses multiple mathematical methods it only rely on accelerometers) and interoperability give good marks in all the evaluated criteria.

Second place was to CMU team who achieved the best marks in accuracy, shown in Fig. 13 (confusion matrix is shown in Table 5), delay (drawn with USS team) and user acceptance criteria. However their localization system, used to improve the accuracy of this solution, was composed by 47 radio beacons and its installation lasted one hour, furthermore another member of the team spent 27 minutes to calibrate the rec-

---

[5]http://www.mobilerobots.com/researchrobots/p3at.aspx
[6]http://store.irobot.com/family/index.jsp
[7]http://en.wikipedia.org/wiki/Kinect
[8]http://www.zephyr-technology.com/bioharness-bt

[9]http://span.ece.utah.edu/radio-tomographic-imaging
[10]http://research.microsoft.com/en-us/um/cambridge/projects/sensecam/

Table 4

Best performance result

Team	Accuracy	Delay	Installation	User acceptance	Interoperability	Final score
USS	4.33	9	10	7.47	7.63	7.3945
CMU	7.17	9	0	7.93	6.15	6.4975
CUJ	2.59	2	0	5.6	5.09	3.5235
DCU	0	0	10	5.2	1.25	2.9875

Table 5

Confusion matrix from CMU best performance

	Stand	Walk	Sit	Bend	Cycle	Fall	Lie
Stand	36.71%	58.23%	0.00%	0.00%	2.53%	2.53%	0.00%
Walk	0.00%	91.59%	0.00%	0.93%	1.87%	5.61%	0.00%
Sit	4.05%	4.05%	31.08%	33.78%	5.41%	21.62%	0.00%
Bend	0.00%	2.94%	0.00%	76.47%	11.76%	8.82%	0.00%
Cycle	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%
Fall	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%
Lie	0.00%	0.93%	1.87%	4.67%	0.00%	8.41%	84.11%

Fig. 12. Pictures from CIAMI living room and garden.

ognizing system based on the accelerometer of chest strap and the mobile phone. This caused the minimum mark in installation time.

Third place was to UCJ team, whose robot was not able to achieve good marks in accuracy and delay crite-ria. Furthermore, installation time took one hour (lead-ing to mark 0 in installation complexity metric) due to the time required to prepare the software and the bat-teries. The UCJ solution seems an early prototype but some important lessons were learned:

– Floor surface can be a problem: Roomba was not able to move through the artificial grass in the in-door garden (Fig. 12 up). Although more pow-erful robot like Pioneer probably can move cor-rectly, other kind of surfaces and steps can create some problems to it.
– The robot is not prepared to follow users through narrow spaces such as between the table and sofa shown in Fig. 12 down.
– A robot can bother the user if it obstructs narrow spaces such as corridors so it must predict user movements in order to avoid obstruct users move-ments.

6. Conclusions

In the indoor localization track most teams used signal strength and presence sensors. This is proba-bly an indication that there is a form of convergence towards the approaches that better fit localization for AAL. However, there are still many differences from the point of view of technologies used to implement the localization methods. As compared to the EvAAL 2011 results, the accuracy of the systems has been im-proved significantly, which is an indication that solu-tions for localization are still evolving.
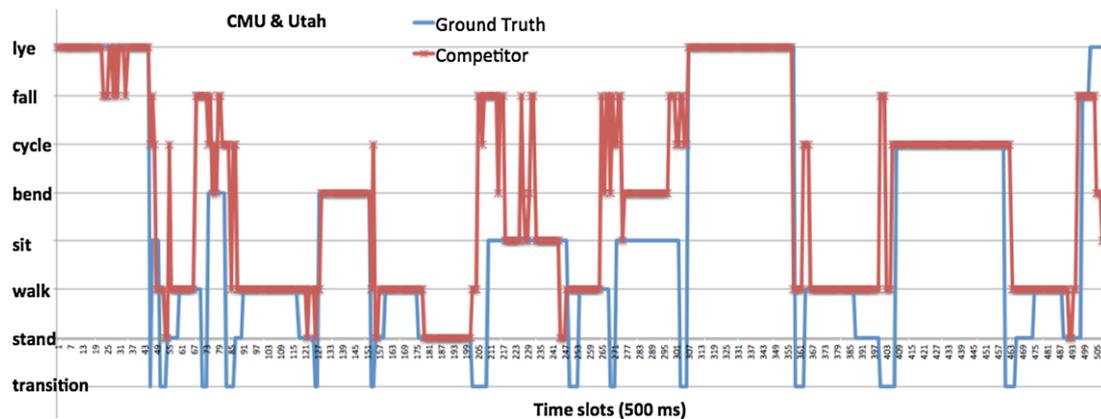
Fig. 13. CMU team accuracy results.

On the other hand, the solutions for activity recognition appears less mature, as witnessed by the diversity of approaches and even performance achieved. Possible improvements in this competition may regard the kind of activities (some more complex activities can be taken into account) and the possibility to make available the environmental sensors in the living labs to the competitors. These results suggest that there is still space for other editions of these competitions in the future, and encourage us to continue in this track for the next year.

Acknowledgements

References

[1] J. Augusto, M. Huch, A. Kameas, J. Maitland, P. McCullagh, J. Roberts, A. Sixsmith, and R. Wichert, eds, *Handbook of Ambient Assisted Living. Technology for Healthcare, Rehabilitation and Well-being*, IOS Press, 2012.

[2] Ambient Assisted Living roadmap, 2009. [Online]. Available: http://www.aaliance.eu/public/documents/aaliance-road map/.

[3] AAL Forum – Ambient Assisted Living Forum. [Online]. Available: http://www.aalforum.eu/.

[4] K. Connelly, K. Siek, I. Mulder, S. Neely, G. Stevenson, and C. Kray, Evaluating pervasive and ubiquitous systems, *IEEE Pervasive Computing* **7**(3) (2008), 85–88.

[5] EvAAL competition. [Online]. Available: http://evaal.aaloa.org.

[6] AALOA association. [Online]. Available: http://aaloa.org.

[7] Smarthouse Living Lab UPM. [Online]. Available: http://smarthouse.lst.tfo.upm.es/.

[8] A. Martínez, M.A. Llorente, and J.P. Lázaro, Ciami living lab: an economically sustainable technological tool for open innovation, in: *I ENoLL Living Lab Summer School*, 2010. [Online]. Available: http://www.ictusagelab.org/ecoleLL/sites/default/files/Position_paper_Summer_School_TSB_2010_v2.doc.

[9] Trading Agent Competition, 2010, http://www.sics.se/tac/.

[10] DARPA Grand Challenge, 2007, http://www.darpa.mil/grandchallenge/index.asp.

[11] N.V. Shilov and K. Yi, Engaging students with theory through acm collegiate programming contest, *Commun. ACM* **45**(9) (2002), 98–101. [Online]. Available: http://doi.acm.org/10.1145/567498.567506.

[12] M. Mukund, International olympiad in informatics, in: *Resonance*, Vol. 7, 2002, pp. 102–107 doi:10.1007/BF02836177. [Online]. Available: http://dx.doi.org/10.1007/BF02836177

[13] C. Wolf, J. Mille, L. Lombardi, O. Celiktutan, M. Jiu, M. Baccouche, E. Dellandrea, C.-E. Bichot, C. Garcia, and B. Sankur, The liris human activities dataset and the icpr 2012 human activities recognition and localization competition, LIRIS Laboratory, Tech. Rep. RR-LIRIS-2012-004, March 2012.

[14] D. Roggen, Collecting complex activity data sets in highly rich networked environments, in: *Seventh International Conference on Networked Sensing Systems (INSS10)*, 2010.

[15] H. Sagha, S.T. Digumarti, J. del R. Millán, R. Chavarriaga, A. Calatroni, D. Roggen, and G. Tröster, Benchmarking classification techniques using the opportunity human activity dataset, in: *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference*, Oct. 2011, pp. 36–40.

[16] D. Roggen, Opportunity activity recognition, UCI Machine Learning Repository, June, 2012. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/OPPORTUNITY+Activity+Recognition

[17] N. Kawaguchi, N. Ogawa, Y. Iwasaki, K. Kaji, T. Terada, K. Murao, S. Inoue, Y. Kawahara, Y. Sumi, and N. Nishio, Hasc challenge: Gathering large scale human activity corpus for the real-world activity understandings, in: *ACM AH 2011*, 2011.

[18] N. Ogawa, K. Kaji, and N. Kawaguchi, Effects of number of subjects on activity recognition – findings from

hasc2010corpus, in: *Proc. of 1st International Workshop on Frontiers in Activity Recognition using Pervasive Sensing (IW-FAR2011)*, 2011, pp. 48–51.

[19] L. Vitali, G. Hassan, K. Latifur R., and J. Roozbeh, A mining technique using n-grams and motion transcripts for body sensor network data repository, in: *Wireless Health'10*, 2010.

[20] L. Vitali and J. Roozbeh, Power aware wireless data collection for bsn data repositories, in: *International Conference on Body Sensor Networks*, 2011.

[21] H. Liu, H. Darabi, P. Banerjee, and J. Liu, Survey of wireless indoor positioning techniques and systems, *IEEE Transactions on Systems, Man, and Cybernetics, Part C* (2007), 1067–1080.

[22] S. Hanke, UniversAAL – an open and consolidated AAL platform, in: *Ambient Assisted Living. 4. AAL-Kongress 2011*, 2011, pp. 127–140.

[23] A.R. Jiménez Ruiz, F. Seco Granja, J.C. Prieto Honorato, and J.I. Guevara Rosas, Accurate pedestrian indoor navigation by tightly coupling foot-mounted imu and rfid measurements, *IEEE T. Instrumentation and Measurement* **61**(1) (2012), 178–189.

[24] S. Wagner, M. Handte, M.A.Z. Zamalloa, and P.J. Marrón, On optimal tag placement for indoor localization, in: *PerCom*, 2012, pp. 162–170.

[25] M. Cypriani, P. Canalda, and F. Spies, Owlps: A self-calibarted fingerpringing-based Wi-Fi positioning system, in: *Communications in Computer and Information Science*, Vol. 309, 2012, pp. 36–51.

[26] J. Wilson and N. Patwari, See-through walls: Motion tracking using variance-based radio tomography networks, *IEEE Trans. Mob. Comput.* **10**(5) (2011), 612–621.

[27] A.V. Medina, J.A. Gómez, O. Rivera, E. Dorronzoro, and M. Merino, Fingerprint indoor position system based on openmac, in: *WINSYS*, 2011, pp. 47–52.

[28] S. Knauth, L. Kaufmann, R.k. Christian Jost, and A. Klapproth, The iloc ultrasound indoor localization system at the EvAAL 2011 competition, in: *Communications in Computer and Information Science*, Vol. 309, 2012, pp. 52–64.

[29] N.M. Boers, D. Chodos, J. Huang, P. Gburzynski, I. Nikolaidis, and E. Stroulia, The smart condo: Visualizing independent living environments in a virtual world, in: *PervasiveHealth*, 2009, pp. 1–8.

[30] N. Noury, A. Fleury, P. Rumeau, A.K. Bourke, G. ÓLaighin, V. Rialle, and J. Lundy, Fall detection – principles and methods, in: *Proc. of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS 2007)*, 2007, pp. 1663–1666. [Online]. Available: http://subversion.assembla.com/svn/fallD/trunk/Stage-I/Literature/Fall%20detection.pdf.

[31] M. Nergui, Y. Yoshida, J. Gonzalez, Y. Koike, M. Sekine, and Y. Wenwei, Human motion tracking and measurement by a mobile robot, in: *The International Conference on Intelligent Unmanned Systems 2011 – ICIUS2011*, 2011.

[32] J.-H. Hong, J. Ramos, and A.K. Dey, Understanding physiological responses to stressors during physical activity, in: *Proc. of the 2012 ACM Conference on Ubiquitous Computing*, Ser. UbiComp '12. ACM, New York, NY, USA, 2012, pp. 270–279. [Online]. Available: http://doi.acm.org/10.1145/2370216.2370260.

[33] N. Li, M. Crane, and H.J. Ruskin, Automatically detecting significant events on sensecam, *Ercim News* **87** (2011), 30–31.

[34] L.G. Abril, F.J. Cuberos, F. Velasco, and J.A. Ortega, Ameva: An autonomous discretization algorithm. *Expert Syst. Appl.* **36**(3) (2009), 5327–5332. [Online]. Available: http://dblp.uni-trier.de/db/journals/eswa/eswa36.html#AbrilCVO09.