Competitive Live Evaluation of Activity-recognition Systems

Authors:

Hristijan Gjoreski, Simon Kozina, Matjaž Gams, Mitja Luštrek
Jožef Stefan Institute, Slovenia
Jožef Stefan International Postgraduate School, Slovenia

Juan Antonio Álvarez-García
University of Seville, Spain

Jin-Hyuk Hong, Julian Ramos, Anind K. Dey
Carnegie Mellon University, USA

Maurizio Bocca, Neal Patwari
University of Utah, USA

Abstract

In order to ensure the validity and usability of activity recognition approaches, an agreement on a set of standard evaluation methods is needed. Due to the diversity of the sensors and other hardware employed, designing and accepting standard tests is a difficult task. This article presents an initiative to evaluate activity recognition systems: a living-lab evaluation established through an annual competition − EvAAL-AR (Evaluating Ambient Assisted Living Systems through Competitive Benchmarking − Activity Recognition). In the competition, each team brings their own activity-recognition system, which is evaluated live on the same activity scenario performed by an actor. The evaluation criteria attempt to capture the practical usability: recognition accuracy, user acceptance, recognition delay, installation complexity, and interoperability with ambient assisted living systems. The article also presents the competing systems with emphasis on two best-performing ones: (i) a system that achieved the best recognition accuracy, and (ii) a system that was evaluated as the best overall. Finally, the article presents lessons learned from the competition and ideas for future development of the competition and of the activity recognition field in general.

1. Introduction

Ambient Assisted Living (AAL) is a research area that uses technology to improve the quality of life of the elderly, by increasing their autonomy in daily activities, and by enabling them to feel secure, protected and supported. It is motivated by the ageing of the population: projections for the developed world show that by 2050, every elderly person will be supported by only two persons of working age, as opposed to four in 2012 [1]. AAL solutions employ various sensors, both wearable (such as accelerometers) and ambient (such as cameras, or sensors built into furniture and home appliances). They perform a wide range of functions, such as health monitoring, help with daily activities, and communication with family, friends and caregivers.

In order to ensure the validity and usability of AAL approaches, leading to eventual real-life application, researchers should agree on a set of standard evaluation methods. However, this is a difficult task because of the diversity of both the functions performed by AAL solutions, as well as the sensors and other hardware they employ. Additionally, the evaluation tests should be as close to real life as possible, since real life poses different problems than laboratory tests. A function that has emerged as central to many AAL solutions is Activity Recognition (AR), mainly because the ability to understand the user's situation and context is key for real-life usability. This article thus presents an initiative to evaluate AR Systems (ARS) on recognizing basic activities in a close approximation of a real-life environment, i.e., living lab.

The first ARS were evaluated on datasets specific to each system, recorded in laboratory settings, and they generally achieved high accuracy. However, since each research group evaluated their system using their own dataset, comparison between different ARS was almost impossible. The first improvement towards standardized test was achieved when the first benchmark datasets were developed: OPPORTUNITY dataset [2], HASC activity recognition corpus [3], HARL competition datasets [4], datasets in the AmI Repository [5], etc. These datasets finally enabled researchers to compare different ARS. Such an evaluation approach is well established in many research areas related to artificial intelligence: examples include the UCI Machine Learning Repository, TREC (Text REtrieval Conferences) that feature competitions in various disciplines, etc. However, for ARS this approach is not sufficient for two reasons. First, the comparison is limited to systems that use the same sensor configuration (type and placement of sensors, etc.) as the one used while recording the dataset. Most of the benchmark datasets are recorded with wearable inertial sensors and can thus only be used to compare ARS that use inertial sensors. And second, benchmark datasets allow only the comparison of the data-processing parts of ARS. This is a severe limitation because it is often the data-acquisition part (the sensors) that limits the systems' reliability and acceptability, and thus their real-life usability.

In this article, we present a living-lab evaluation of ARS established through an annual competition − EvAAL-AR (Evaluating AAL Systems through Competitive Benchmarking − Activity Recognition). In the competition, each team is required to bring their own ARS, which is evaluated using criteria capturing its practical usability: recognition accuracy, user acceptance, recognition delay, installation complexity, and interoperability with AAL systems. The performance of each competing system is evaluated live on a predefined activity scenario performed by an actor. The article also describes the competing systems with an emphasis on two best-performing ones: the system that achieved the best recognition accuracy, and the system that was evaluated as the best overall. Finally, we present the lessons learned during the competition, discussion on future development of ARS and the competition.

2. EvAAL Activity Recognition Competition

The EvAAL competition was conceived in 2010, as one of the main objectives of the universAAL project (http://universaal.org/index.php/en/). It was designed around the grand challenge of evaluating complete AAL systems, with a long-term roadmap that starts with the evaluation of simple components and building blocks in the first phase, and continues with the evaluation of aggregated components, services and even platforms in the second phase. It started in 2011 as a single-track localization competition [6], evaluating only systems for indoor localization, and later evolved into a two-track competition by including the AR track (EvAAL-AR) in 2012 ('12) and 2013 ('13) [7].

The main objective of the EvAAL-AR competition is to evaluate ARS intended to be used by the elderly in real life, therefore the performance of each competing ARS is evaluated live in a living lab. The '12 and '13 EvAAL-AR competitions were held in the CIAmI Living Lab in Valencia, Spain, shown in Figure 1. There was no limitation to the number and type of devices comprising the ARS brought by the competitors. The only constraint was the compatibility with the physical limitations of the living lab.

Figure 1. CIAMI Living Lab, Valencia, Spain.

The following seven activities had to be recognized: lying, sitting, standing, walking, bending, cycling on a stationary bike, and falling. Most of them were selected because they are common in daily life and thus recognizing them is the starting point for AR. Together with the localization of the user (tackled by the other EvAAL track), they provide the context for smart control of home automation, and paint a broad picture of an elderly's lifestyle, including the very important level of physical activity. Falling was included because it is the main cause of injury among the elderly, while cycling is a recommended type of exercise for older people. All the activities were included in a scenario that lasted approximately 5 minutes, performed by an actor. The scenario represented a

simulation of a part of the day of an elderly (watching TV, working in the kitchen, bathroom activities, sleeping) and was repeated twice, with the better run by each competitor counting towards the final score. In order to get approximately the same ground truth for all the competitors, audio cues were played from a file to signal the actor as to which activity should be performed three seconds in advance, giving the actor the time to prepare for it. An evaluator who followed the actor refined the ground truth by marking the precise time-stamps of the activities with a custom smartphone application.

The scoring of the ARS was overseen by an evaluation committee (EC) according to the following five criteria. The numbers in the brackets represent the weights by which each criterion was considered in the final score in the '12 competition:

- (25%) Recognition accuracy − how accurately the system recognizes the target activities. Even though different types of evaluation metrics can be found in the AR literature (overfill, merge, fragmentation, deletion, accuracy, recall, precision, F-measure, AUC, ROC, etc.) [2][8][9], a single metric was used to rank the competitors – the F-measure. It is calculated as a harmonic mean of the recall and the precision values which are averaged over the seven target activities. This metric is among the most commonly used in AR and is a good estimator of the AR performance. However, all the data (ground truth and recognized activities) is available and can be used to calculate different performance scores using other metrics.

- (20%) Recognition delay – the elapsed time between the time at which the user begins an activity and the time at which the system recognizes it. The maximum allowed delay was 20 seconds, after that the system was evaluated with 0 points. In order to get the maximum score, the competing system was required to have a delay no more than 2 seconds.

- (25%) Installation complexity – how much effort is required to install the ARS in the living lab. It was measured in minutes of work per person needed to complete the installation. The maximum allowed installation time was 60 minutes, after that the system was evaluated with 0 points. In order to get the maximum score, the competing system was required to have no more than 10 minutes of installation time. This and the following two parameters were evaluated by the EC.

- (15%) User acceptance − how invasive the ARS is in the user's daily life. Evaluated by the EC using a questionnaire available at: http://evaal.aaloa.org/2013/quest.

- (15%) Interoperability with AAL systems – the metrics used are: the use of open-source solutions, availability of libraries for development, integration with standard protocols. Evaluated by the EC using a questionnaire available at: http://evaal.aaloa.org/2013/quest.

The choice of the criteria and their weights was inspired by the evaluation criteria used in the EvAAL localization competition held in 2011. The choice was also supported by the following assumptions: (1) the **accuracy** is the best way to determine the operational quality of the system; (2) the **delay** is an indication of whether the system works in real-time; (3) the **installation complexity** is an indication of the adoption barrier for end users, and is used to limit the time that the competitor has to install his system; (4) the **user acceptance** - although it is a subjective criteria, it is an important issue in actual use; and finally (5) the **interoperability** was motivated by being one of the goals of the universAAL project (to create open AAL platforms).

The initial competition setup was based on the organizers' extensive experience with AR, but since such an event had not been organized before, the experience gained in the '12 competition prompted some improvements for the

'13 edition. Firstly, the actor performing the activities wore an elderly simulation kit. This is a specially designed garment to hinder the movements of the actor emulating an elder (the actor wearing the kit is shown in the bottom right panels in Figure 1). Note that having an elder as an actor was not an option because of the risky fall event and the physical burden of repeating the activities for each team. Secondly, the weight for accuracy was increased (from 25% to 35%) and weights for installation complexity and interoperability with AAL systems reduced (from 25% to 20% and 15% to 10%, respectively). These changes were introduced to better reflect the current state of AR development, since the main goal is still to achieve adequate accuracy in real life, while the installation complexity and interoperability will come to the fore once the area matures. Thirdly, several kinds of falls were considered, including backward, forward and lateral falls.

During the competition, each competing system recorded a dataset using its own sensor configuration. These datasets were labeled with the appropriate activity and can be used for offline comparison of different algorithms and data-processing techniques. Eight labeled datasets are available at the competition's web-site (http://evaal.aaloa.org/), which include a variety of sensors data: accelerometers, camera images, heart rate, breath rate, etc.

Table 1 shows the scores on the scale of 0–10 for the '12 and '13 editions. Due to the change in the weights of the criteria for the '13 edition, we include the final scores by both years' rules. Even though the system by the Carnegie Mellon and Utah universities (CMU&Utah) experienced some installation problems, it still achieved the best recognition accuracy. However, the system by Jožef Stefan Institute (JSI) obtained the highest final score for both years, by achieving not only high accuracy, but also scoring very well on the other criteria.

Table 1. EvAAL-AR '12 and '13 teams and results (score: from 0 to 10)

	Team	Accuracy	Delay	Installation complexity	User Acceptance	Interoper-ability	Overall score '12	Overall score '13
EvAAL-AR '12	Seville'12 (Spain)	4.33	9	10	7.47	7.63	7.39	7.07
	CMU&Utah (USA)	**7.17**	**9**	**0**	**7.93**	**6.15**	**6.5**	**6.51**
	Chiba'12 (Japan)	1.44	5	0	5.6	5.09	3.52	3.13
	Dublin (Ireland)	0	0	10	5.2	1.25	2.99	2.67
EvAAL-AR '13	**JSI (Slovenia)**	**6.94**	**10**	**10**	**8.55**	**7.2**	**8.45**	**8.36**
	CNR (Italy)	4.04	10	10	7.04	6.15	7.19	6.94
	Seville'13 (Spain)	4.68	9	10	6.99	5.54	7.05	6.89
	Chiba'13 (Japan)	4.43	10	0	5.44	2.24	4.8	4.86

3. The Competitors

3.1. CMU&Utah System

The CMU&Utah system [10], which participated in the '12 edition only, actually managed to achieve the highest accuracy across both competition years. The system is composed of the BioHarness BT chest strap, a smartphone, an indoor localization system, and a laptop. The BioHarness and the smartphone, carried in the user's pocket, collect 3D accelerations. The indoor localization system estimates the user's indoor location with radio tomographic imaging, measuring the disruption of radio signal strength caused by a moving object that either reflects or absorbs the wireless signal [11]. The data from the three sources are collected on the laptop, which performs the AR.

The CMU&Utah's system uses a hybrid AR methodology [12], with one expert module per activity and a control module. Each expert module uses the data from the BioHarness and outputs the probability of the recognized activity. The control module improves the AR by considering the richer context of the user, including the user's location and the orientation of the smartphone in the pocket. Each new user of the system is required to perform each activity for a short time for the purpose of calibration, using a simple labeling interface on the smartphone. The calibration data is evaluated against a pool of AR models trained on different people, and the best-performing set of models is selected for the new user.

This methodology was selected based on laboratory experiments using data collected for 15 participants. The participants performed a set of tasks, with no explicit instruction or restrictions to ensure realistic movements. Afterwards, three approaches for training the AR models were compared: individual, population and calibration. In the individual approach, for each of the 15 participants half of the data were used to train the AR models, and the other half to test them. In the population approach, for each participant the data from the other 14 participants and half of the participant's data were used for training. The models were then tested on the other half of the participant's data. Finally, for the calibration approach, for each participant the data from the remaining 14 participants were used to train a pool of 14 sets of AR models. Then, half of the participant's data were used for calibration, i.e., finding the best set of models out of the pool, and other half of the data was used for testing.

As shown in Figure 2 (pre-competition part), the calibration approach achieved the highest accuracy of 74% on the test data and was thus used in the '12 competition. At the competition, it achieved similar accuracy of 72% on a previously unknown actor.
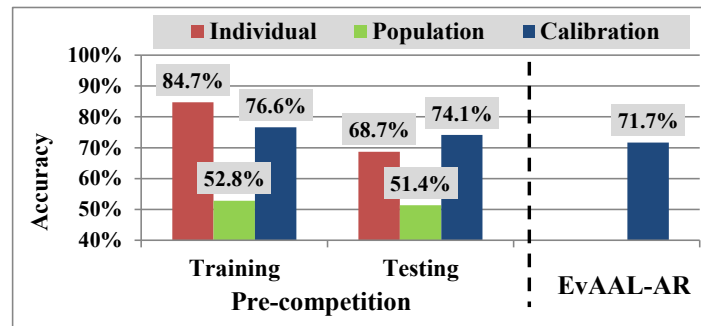

Figure 2. Pre-competition and EvAAL-AR accuracy for the CMU&Utah system.

The main strength of the CMU&Utah system is the personalization of the AR models. Even though the calibration needed for the personalization makes the installation somewhat more complex, it is probably worth it, since the accuracy at the competition was almost the same as in the laboratory experiments. The fact that no pre-defined activity scenario was used during training probably contributed to this result, since the AR models could not overfit to a training scenario. The localization should also increase the accuracy at the expense of installation complexity. Since it failed to work in the living lab due to wireless interference with the equipment in the environment, it was not used at the competition, but the modularity of the system made it possible to compete without it and still achieve the best recognition accuracy. However, the attempts to make the localization work resulted in a low installation complexity score and thus prevented the victory. This shows that sophisticated AR methods are not necessarily sufficient for success in the field.

There are several improvements considered for future work: the encryption of the localization data to avoid the interference with other equipment, improving the accuracy by including the heart rate and breath rate outputs from the BioHarness chest strap, and finally smartphone implementation of the system so it can also be used outdoors.

3.2. JSI System

The JSI system [13], competing in the '13 edition only, actually achieved the highest ranking overall across both competition years. It is composed of two accelerometers sewn into clothing, placed on the user's abdomen and thigh. The placement was chosen as a trade-off between the physical intrusiveness and the accuracy in preliminary tests [13]. The accelerometers use Bluetooth to transmit data to a laptop, where AR is performed. The software architecture of the system deals with AR and fall detection (FD) separately. First the FD module checks whether a fall has occurred, and if not, the AR module outputs the activity.

The FD module detects a fall pattern: a decrease in acceleration (falling) followed by an increase (impact) [14]. The minimum and the maximum acceleration within a one-second window are measured, and the difference between them has to exceeded 10 m/s^2. Following a fall pattern, if the abdomen sensor is in the horizontal orientation, the fall event is confirmed.

In the AR module, a number of features are first computed from the acceleration data. The activities are then recognized by a three-level scheme [13]. On the first level, a classifier was trained only to distinguish the cycling activity from the others. If the activity is not classified as cycling, the feature vector is passed to the second level, where the postures (sitting, lying, bending, and the upright posture) are recognized by rules. If the recognized activity is the upright posture, a classifier on the third level is used to distinguish between standing and walking.

To train the AR classifiers and to evaluate and tune the methodology, a complex 90-minute scenario was recorded by 10 participants. The AR model was evaluated by the leave-one-person-out population approach, so the model for each participant was constructed from the data from the remaining 9 participants. The data from all 10 participants were used for training the AR model used in the competition.

Figure 3 shows the performance of the JSI system evaluated prior to the EvAAL-AR competition, compared to the results from the competition. The JSI system performed much less accurately in the competition than in the laboratory experiments beforehand (overall F-measure of 69% vs. 94%). The recognition accuracy of the JSI system in the pre-competition experiments was very good because the tested activity scenario was the same for all the

participants and was performed in a fairly controlled fashion. Because of that, somewhat worse results were expected at the competition. However, the accuracy at the competition was much lower, mainly because the actor at the competition was strongly leaning back when sitting, so that the angle of the abdomen sensor was very low. In the training data, the participants were sitting much more uprightly, so most of the sitting at the competition was misrecognized as lying. This clearly shows a need for personalization of the ARS, e.g., by calibrating the system and adjusting the models to the current conditions (sensor orientation angles).
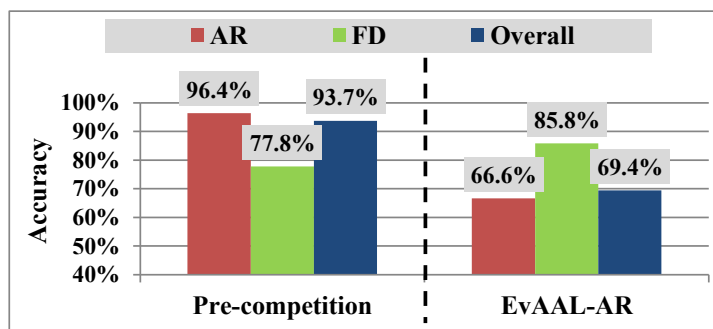

Figure 3. Pre-competition and EvAAL-AR accuracy for the JSI system.

Even though the JSI system performed best at the competition, the experience suggested several aspects that would benefit from improvement: to reduce the number of sensors to one, to use smaller and more robust sensors, to improve the interoperability with other systems and applications (providing API and services to other AAL applications), and smartphone implementation for outdoor AR. The JSI team is also considering personalization along the lines of CMU&Utah system.

3.3. Other Systems

The **Chiba** team participated in '12 and '13 editions. In 2012 it used a Roomba robot (http://store.irobot.com/family/index.jsp) with a laptop and two Kinect 3D scanners (http://en.wikipedia.org/wiki/Kinect), the first one to avoid obstacles when following the actor and the second one to recognize his activities. Roomba proved underpowered to carry the equipment, so in 2013 a Pioneer 3-AT robot (http://www.mobilerobots.com/researchrobots/p3at.aspx) was used instead, and only one Kinect did both tasks. However, this robot required mapping the living-lab in advance, resulting in low score for installation complexity criteria. Additionally, the robot could not pass through narrow spaces and could not get to the garden due to a step. This and the inability to cope with the bright lighting in the living lab prevented a high accuracy score.

The **Seville** team also participated in both editions with a smartphone placed on the right hip. The embedded accelerometer was used to recognize the user's activities. The phone was fixed with a belt to ensure stable placement. This detracted from the user acceptance, although it was still relatively high, while not being enough to achieve high accuracy. Low delay and installation complexity were sufficient for the victory in 2012, but not in 2013.

The **CNR** team used a smartphone placed in the front trousers pocket. In addition, three nodes with embedded radio transmitter/receivers were placed on the chest and both ankles, and another one on the stationary bike. The accelerometer embedded in the phone was the main sensor for the AR, while the received signal strength between

the radio nodes provided some localization information, mainly the distance to the bike. As in the case of the Seville team, one accelerometer was not enough for accurate AR.

The **Dublin** team used a SenseCam camera (http://research.microsoft.com/en-us/um/cambridge/projects/sensecam/) hanging around the neck. The sampling frequency of the camera was not sufficient to detect changes in the activities of the user, and the analysis of the data was performed offline. These issues resulted in low scores for the accuracy and recognition delay.

4. Discussion

AR is an essential part of ambient intelligence and as such needs standard evaluation methods. We described the EvAAL-AR competition, whose ambition is to become a "gold standard" of AR quality. In comparison to benchmarking on datasets, the competition evaluates the performance of ARS live (in real-time) and does not limit the competing systems to a predefined sensor configuration and data, since each system captures its own data. Furthermore, it does not allow tuning to a particular dataset, which may happen with a benchmark dataset. This was observed in the JSI's pre-competition and competition results, where a large drop in performance occurred. Finally, the competition allows evaluating ARS by criteria related to practical usability, not only by their recognition performance.

The EvAAL-AR competition has been an interesting experience both for the competing teams and the organizers, despite all of them having substantial experience with AR. Some lessons that have been learned:

- Since participating in the competition requires an investment in time and money, only teams that were confident in the quality of their ARS participated. However, they encountered numerous unexpected problems, such as those by the CMU&Utah and JSI teams, as well as equipment being damaged in transit and sensors falling off the actor. This is evidence that most of ARS developed in research laboratories need modifications for practical use.

- Simultaneously detecting activities of daily living and falls presented quite a challenge for the competing systems, suggesting that AR and FD are being developed in isolation instead of as parts of a single system.

- While the evaluation scenario was short and relatively simple, the impression of people involved – the authors as well as others – was that it is a decent indicator or real-life performance. The elderly simulation kit helped emulate the movements of a >65 years old person. A longer and more complex evaluation (multiple days of real life) would be preferable, but too difficult and expensive to organize.

- The competition was a good environment for discussion and first-hand observation of different ARS, providing valuable feedback to improve both the ARS and the competition. The competition improves with each next year and should eventually evaluate all the necessary aspects of creating a complete ARS.

- The best-performing systems recognized basic activities adequately. Such activities are building blocks of complex activities (e.g., preparing a meal) whose recognition is often more central to the tasks of AAL systems, and it is probably time to take the evaluation a step forward and tackle complex activities directly.

The main objectives for the next editions of the EvAAL-AR are: (i) to further the goal of establishing the competition as the "gold standard" for measuring quality of ARS, (ii) to increase the number of competitors, and (iii) to upgrade the test scenario with more complex activities in less controlled environments. To achieve the first two objectives, we are working to organize it in conjunction with a high-level conference that attracts more interested researchers than the current competition event, include more AR experts worldwide, and promote it in the AR community. Attracting vendors and increasing the value of the award is another possibility. As to the last objective, we are considering to include:

- A simulated shower or toilet visit. These two events are of particular interest because on one hand those are the places where critical falls often occur, and on the other hand are technologically challenging because of the privacy, safety and sensor wearability issues.

- More challenging FD by including some intentional fall-like events that may trigger a false alarm and possibly additional fall types.

- Recognition of high-level activities of daily life, e.g., having a conversation, watching TV, reading a newspaper, cooking, etc., which are still a challenging task in the pervasive computing [15].

- Outdoor evaluation. Most of the ARS assume that the user is inside a controlled environment. However, outdoor monitoring is an important aspect and a challenging task, especially for physically active users.

In order to improve the evaluation, we propose:

- Further refining the evaluation criteria, particularly the recognition delay and the recognition accuracy. First, while it is important to recognize activities in a timely manner, a few seconds are usually not very important. Second, a more thorough analysis of the recognition accuracy is needed, e.g., including per-activity analysis – some activities are more important than others, e.g., falling vs. standing.

- Improving the questionnaires. Although the technical committee members, who designed the questionnaires, are experts in AR (http://evaal.aaloa.org/2013/committees2013), some questions were not objective enough and require further refinement. Experts from other areas, such as usability, should also be involved.

Acknowledgments

References

[1]   United Nations, Department of Economic and Social Affairs, Population Division, "Population aging and development 2012", United Nations, 2012.

[2]   R. Chavarriaga, H. Sagha, A. Calatroni, S. Digumarti, G. Tröster, J. del R. Millán, D. Roggen. "The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition", Pattern Recognition Letters, 2013.

[3] N. Kawaguchi, N. Ogawa, Y. Iwasaki, K. Kaji, T. Terada, K. Murao, S. Inoue, S., Y. Kawahara, Y. Sumi, and N. Nishio, "HASC Challenge: Gathering Large Scale Human Activity Corpus for the Real-World Activity Understandings," *Proc. ACM AH*, 2011, pp.27:1-27:5.

[4] C. Wolf, J. Mille, E. 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," *Technical Report* RR-LIRIS-2012-004, LIRIS Laboratory, 2012.

[5] B. Kaluža, S. Kozina, and M. Luštrek. The activity recognition repository: towards competitive benchmarking in ambient intelligence. *Proc. Activity Context Representation Workshop, AAAI,* 2012.

[6] P. Barsocchi, S. Chessa, F. Furfari and F. Potorti, "Evaluating Ambient Assisted Living Solutions: The Localization Competition", *Pervasive Computing*, IEEE, vol. 12, no. 4, 2013,pp. 72–79.

[7] J. A. Álvarez-García, P. Barsocchi, S. Chessa, and D. Salvi, "Evaluation of localization and activity recognition systems for ambient assisted living: The experience of the 2012 EvAAL competition," *Journal of Ambient Intelligence and Smart Environments*, 2013, pp. 119–132.

[8] J. A. Ward, P. Lukowicz, H. W. Gellersen, "Performance metrics for activity recognition". *ACM Transactions on Intelligent Systems and Technology,* 2011, 2(1), 6.

[9] A. Reiss and D. Stricker, "Creating and benchmarking a new dataset for physical activity monitoring". *In Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments*. 2012 (p. 40). ACM.

[10] J.-H. Hong, J. Ramos, C. Shin, and A. Dey "An activity recognition system for ambient assisted living environments," *Evaluating AAL Systems Through Competitive Benchmarking, Communications in Computer and Information Science*, vol. 362, 2013, pp. 148–158.

[11] M. Bocca, O. Kaltiokallio, and N. Patwari, "Radio Tomographic Imaging for Ambient Assisted Living," *Evaluating AAL Systems Through Competitive Benchmarking*, 2013, vol. 362, pp. 108–130.

[12] J.-H. Hong, J.-K. Min, U.-K. Cho and S.-B. Cho, "Fingerprint classification using one-vs-all support vector machines dynamically ordered with naive Bayes classifiers," *Pattern Recognition*, 2008, vol. 41, no. 2, pp. 662–671.

[13] S. Kozina, H. Gjoreski, M. Gams, and M. Luštrek, "Efficient Activity Recognition and Fall Detection Using Accelerometers," *Evaluating AAL Systems Through Competitive Benchmarking, Communications in Computer and Information Science*, Springer Berlin Heidelberg, 2013, vol. 386, pp. 13–23.

[14] H. Gjoreski, M. Luštrek, and M. Gams, "Context-Based Fall Detection using Inertial and Location Sensors," *Proc. Int. Joint Conf. Ambient Intelligence*, LNCS 7683, Springer, 2012, pp. 1–16.

[15] K. Eunju, S. Helal, and D. Cook, "Human activity recognition and pattern discovery," *Pervasive Computing*, IEEE, 2010, pp. 48–53.

Author Bios

Hristijan Gjoreski is a researcher at the Department of Intelligent Systems at Jožef Stefan Institute. His research interests include context-based reasoning, wearable computing and ambient intelligence. He holds a M.Sc. degree in Information and Communication Technologies from the Jožef Stefan International Postgraduate School. Contact him at hristijan.gjoreski@ijs.si.

Simon Kozina is a researcher at the Department of Intelligent Systems at Jožef Stefan Institute. His main research interest is ambient intelligence. He holds a B.Sc. degree in Computer and Information Science from the University of Ljubljana. Contact him at simon.kozina@ijs.si.

Matjaž Gams is the head of the Department of Intelligent systems at Jožef Stefan Institute, and professor at the University of Ljubljana and Jožef Stefan Postgraduate School. His research interests include ambient intelligence, machine learning, agents, hybrid learning and reasoning. He holds a Ph.D. degree in Computer and Information Science from the University of Ljubljana. Contact him at matjaz.gams@ijs.si.

Mitja Luštrek is the head of the Ambient Intelligence Group at the Department of Intelligent Systems at Jožef Stefan Institute. His main research interest is ambient intelligence, particularly the analysis of human behavior using sensor data. He holds a Ph.D. degree in Computer and Information Science from the University of Ljubljana. Contact him at mitja.lustrek@ijs.si.

Juan Antonio Álvarez-García is an Associate Professor at the Department of Languages and Computer Systems at the University of Seville. His research focuses in human mobility prediction and activity recognition. He holds a Ph.D. degree in Software Engineering from the University of Seville. Contact him at jaalvarez@us.es.
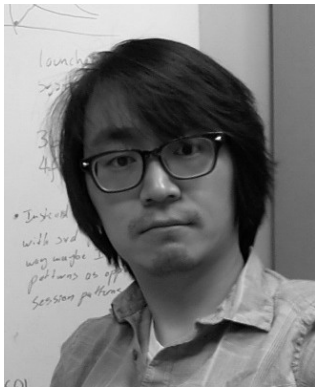
Jin-Hyuk Hong is a systems scientist at Human-Computer Interaction Institute at Carnegie Mellon University. His research interests include mobile computing, context awareness, pattern recognition, and user interface, particularly focusing on the understanding of human behaviors. He holds a Ph.D. degree in Computer Science from Yonsei University. Contact him at hjinh7@gmail.com.

Julian Ramos is a Research programmer at the Human-Computer Interaction Institute at Carnegie Mellon University. His research interests focus on novel applications of Machine Learning in human-computer interaction on topics like Activity Recognition, Stress Recognition and Human Routine Behavior modeling and detection. He holds a bachelor's degree in Mechatronics Engineering. Contact him at ingenia@andrew.cmu.edu.

Anind K Dey is an Associate Professor in the Human-Computer Interaction Institute at Carnegie Mellon University. His research interests include context-awareness, infrastructures to support ubiquitous computing, and the use of embedded and mobile sensors to opportunistically infer human behavior. He holds a PhD degree in Computer Science from Georgia Tech. Contact him at anind@cs.cmu.edu.

Maurizio Bocca is a post doc researcher at the Sensing and Processing Across Networks (SPAN) Lab at the University of Utah. His research interests include RF sensor networks for indoor device-free localization, context awareness and elder care. He holds a Ph.D. degree in Electrical Engineering from Aalto University. Contact him at maurizio.bocca@utah.edu.

Neal Patwari is an Associate Professor of Electrical and Computer Engineering at the University of Utah, and the Director of Research at Xandem Technology. His research interests are in radio channel signal processing, in which radio channel measurements are used to benefit security, networking, and localization applications. He holds a Ph.D. degree in Electrical Engineering from the University of Michigan, Ann Arbor. Contact him at npatwari@ece.utah.edu.