

Human Activity Recognition: from Controlled Lab Experiments to Competitive Live Evaluation

Hristijan Gjoreski, Matjaž Gams, Mitja Luštrek
Department of Intelligent Systems
Jožef Stefan Institute
Ljubljana, Slovenia
{hristijan.gjoreski, matjaz.gams, mitja.lustrek}@ijs.si

Abstract—Human activity recognition is a basic building block in numerous healthcare systems, mainly because the ability to understand the user’s situation and context. This paper presents a solution to the general problem with evaluation of human activity recognition systems, i.e., an activity recognition system may perform perfectly in controlled lab experiments, but significantly worse once applied to more realistic conditions. The solution is presented through the practical experience gained with the creation of our RAReFall activity recognition system. Although the system was awarded the first place at the EvAAL-AR live competition, the recognition accuracy at the competition was significantly lower compared to the controlled lab experiments performed just before the competition. To overcome the encountered problem we developed an automatic calibration method, which solves the encountered problem by adapting and re-calibrating the accelerometer data in real-time while the user is performing everyday activities. The method increased the overall accuracy for 8 percentage points and for 51 percentage points for the sitting activity.

Keywords— *Activity recognition; fall detection; data mining; accelerometer; competition; calibration;*

I. INTRODUCTION AND BACKGROUND

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 [1]. AAL solutions employ various sensors, both wearable (such as accelerometers) and ambient (such as cameras, or sensors built into furniture and home appliances) and perform a wide range of functions in order to assist the users in their everyday life. A function that has emerged as central to many such systems is human Activity Recognition (AR), mainly because the ability to understand the user’s situation and context is key for real-life usability. Even though AR is relatively new topic that emerged with the advancements in the sensor technology (especially MEMS), it is fairly popular and exploited mainly because of its practical applications in numerous healthcare systems (fall detection [2], gait anomaly detection [3], energy expenditure estimation [4][5], stress detection [6], behavior monitoring [7], rehabilitation [8]). With the increasing number and diversity of activity-recognition systems (ARS), it is a challenging task to properly evaluate and to compare ARS.

Additionally, the evaluation should be as close to real life as possible, since real life poses different issues than lab tests.

The first ARS were evaluated on datasets specific to each system, recorded in controlled lab settings, and comparison between different ARS was almost impossible. An improvement was achieved with the development of benchmark datasets: OPPORTUNITY [9], HASC [10], AmI Repository [11], etc. These finally enabled researchers to compare different ARS. Even though such an evaluation approach is well established in many research areas related to artificial intelligence (UCI Machine Learning Repository, TREC), it is not sufficient for evaluation of ARS. First, the comparison is limited to systems that use the same sensor configuration as the one used while recording the dataset. 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.

A step towards real-life evaluation of ARS is the EvAAL-AR (Evaluating AAL Systems through Competitive Benchmarking – AR) competition [12][13]. It represents a unique live evaluation of ARS in a living lab. 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.

This paper presents the experience with creating the RAReFall (Real-time Activity Recognition and Fall detection) system. A confirmation to its effectiveness and practical usability is the first place award at the EvAAL-AR competition, achieving the highest score including the previous year’s competition [13]. Despite winning the competition, the achieved recognition accuracy was significantly lower compared to the controlled lab tests, i.e., 78% vs. 99%. This shows the general problem with evaluation of an ARS, i.e., an ARS may perform perfectly in controlled lab experiments but significantly worse once applied to more realistic conditions: different environment, different activity scenario, etc. To overcome this problem, we developed a calibration method that

adapts and re-calibrates the accelerometers data in real-time while the user is performing everyday activities. The method increased the accuracy for 8 percentage points (p.p.) overall, and for 51 p.p. for the sitting activity (which was the most problematic at the competition).

II. RAREFALL SYSTEM

The experience gained in two European projects Confidence [18] and CHIRON [19], enabled us to create the RAReFall (Real-time Activity-Recognition and Fall-Detection) system [20]. The purpose of the RAReFall system is to be used in real life and to be able to recognize activities and detect falls in real-time.

A. Equipment

The implementation of the RAReFall system that was used at the EvAAL-AR competition, is shown in Figure 1. It consists of two accelerometers sewn into elastic clothes, placed on the abdomen and the right thigh. The AR is performed on a laptop using the raw sensors data acquired through Bluetooth.

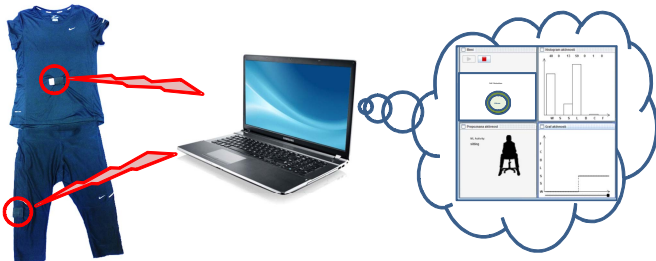


Figure 1. The RAReFall system used at the EvAAL-AR.

As one of the most exploited sensors for AR, 3-axis accelerometers were used in the RAReFall system. The Shimmer sensor platform was used because of its reasonable battery life, compact size, and wireless connectivity. Additionally, a custom sensor firmware was created, so that the sensors can be used to constantly monitor the user during the whole day. The placement of the sensors was chosen as a trade-off between the physical intrusiveness and the performance in preliminary tests [2]. In general though, the user can choose from the 10 locations tested in the preliminary tests and the system would use the appropriate model for AR. To ensure maximum reliability and signal strength, the laptop was equipped with a long-range Bluetooth antenna. In general, though, any kind of Bluetooth device with modest processing capability can be used, therefore a smartphone implementation is considered for future work.

B. Methods

Numerous studies in AR have shown that data mining (DM) approaches can be successfully applied to recognize the basic activities of the user [16][17]. These approaches usually use sliding window to segment the stream of data (e.g., 1 second of data), then to extract numerous features for the particular segment, and finally to recognize the activity by pre-trained classification model.

DM approach was also used as a baseline approach in the RAReFall system (see Figure 2). In the first step, the raw

acceleration data is preprocessed in order to extract multiple features. Because two sensors are used and their data is not received at the same time, first the data is synchronized (aligned). Once the sensor measurements are synchronized, further pre-processing is performed using a band-pass and a low-pass filter. The band-pass filter has two goals: (1) to eliminate the low-frequency acceleration (gravity) that captures information about the orientation of the sensor with respect to the ground and (2) to eliminate the high-frequency signal components generated by non-human motion and high-frequency noise, thus preserving the medium-frequency signal components generated by dynamic human motion. The band-pass filtered data is used for the extraction of features relevant for dynamic activities, such as walking, running and cycling. The low-pass filter has the opposite purpose: to eliminate most of the signal generated by dynamic human motion and preserve the low-frequency component, i.e., gravity. In this way the low-pass filtered data contains the sensor-orientation information, which is relevant for the recognition of the static activities (postures), such as lying, sitting, standing and kneeling. After applying the filters, the acceleration data stream is segmented using sliding window of 2 seconds (1 second overlap). Then, 61 features (a combination of features extracted from the time-domain acceleration signal and frequency domain acceleration signal) are extracted for each sensor. And finally the activity is recognized by pre-trained Random Forest (RF) classification model. The features and the classification algorithm were chosen after performing thorough empirical analysis [2][14].

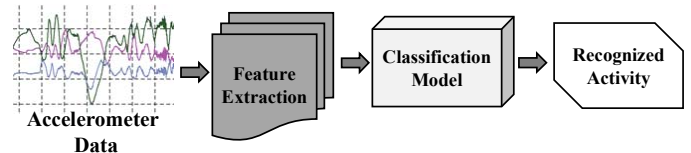


Figure 2. DM approach to activity recognition.

In general, DM models perform and generalize well if the training data fully represents the problem domain. This means that for the AR domain, the training data should include all of the user's postures and activities, i.e., lying should be performed while the user is on the back, on both sides and similar. This is practically impossible task to include all possible postures and types of activities. Because of this, we created a Hybrid model for AR [15], i.e., a model that combines expert rules to recognize postures (static activities), which ensure the behavior of the system is predictable and robust, and classifiers trained with DM algorithms, to recognize dynamic activities, for which the rules are not sufficiently accurate. In particular, the activities are recognized by a three-level scheme, shown in Figure 3 [14] [15]. The structure of the scheme was defined after numerous experiments and empirical analysis of the several AR datasets, available at: <http://dis.ijs.si/ami-repository/>.

Our initial experiments showed that cycling is quite unique activity (periodic movements with the thigh) and that it can easily distinguished with a single DM model. Therefore, on the first level in the Hybrid scheme (see Figure 3), the feature vector is fed into a RF classifier, which is trained to distinguish cycling from the other.

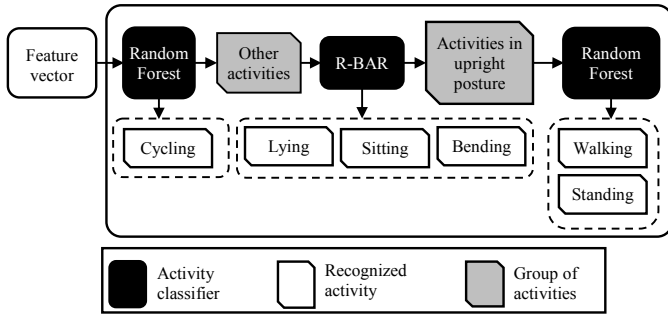


Figure 3. Hybrid DM approach used in RAReFall.

If an activity is not classified as cycling, the feature vector is passed to the second level, where static activities/postures (sitting, lying, bending, and upright posture) are recognized by expert rules (R-BAR). We decided to use expert rules because the static activities, i.e., postures, can be easily described by a human (expert) using the orientation (tilt, inclination) angles of the accelerometer. To calculate the angle between the actual acceleration, i.e., the Earth's gravity (g) for static postures, and some of the axis (e.g., x -axis) we used the following formula (where the values a_x , a_y and a_z represent the actual acceleration vector):

$$\varphi_x = \arccos\left(\frac{a_x}{\sqrt{a_x^2 + a_y^2 + a_z^2}}\right) \quad (1)$$

Figure 4 shows an example of the orientation angles of the chest placed accelerometer when the person is sitting. The calculated angles would be: 180° , 90° and 90° for the x , y and z axis respectively. This kind of reference angles (orientation vectors) are defined for each of the static activities (sitting, lying, bending and upright (standing, walking)) for both accelerometers (torso and thigh). When a new measurement is received, the orientation angles are computed and the orientation vector is constructed and compared to the reference orientation vectors by calculating the similarity/distance to each of them. More details about the chosen similarity metric can be found in our previous work [15]. The one that is the most similar is chosen. If the chosen orientation vector belongs to the activities: sitting, lying, or bending, it is outputted as final; if an upright posture is recognized, the feature vector is fed into the third level, where another RF model distinguishes between standing and walking.

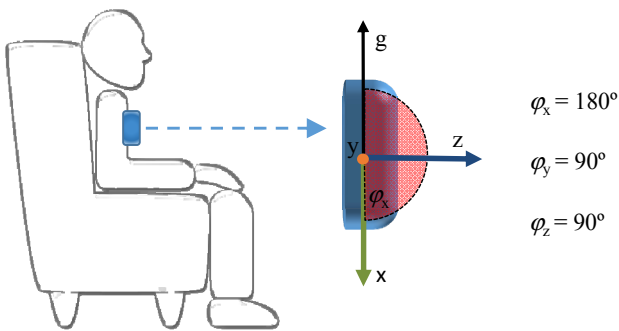


Figure 4. The orientation angles of the chest placed accelerometer when the person is sitting.

Because of its specificity (momentary event which may require attention), the fall detection is tackled separately [14]. Expert rules are used that take into account high accelerations and the recognized horizontal orientation, since falling is often followed by lying. In particular, a threshold was applied to the length of the acceleration vector, which means that the direction of the acceleration was ignored. Then, the minimum and the maximum acceleration within a one-second window were measured. If the difference between the maximum and the minimum exceeded 1 g and the maximum came after the minimum, a fall-like acceleration pattern was detected (see Figure 5). Afterwards, the orientation of the torso accelerometer is checked, and in case of a horizontal orientation a fall was detected.

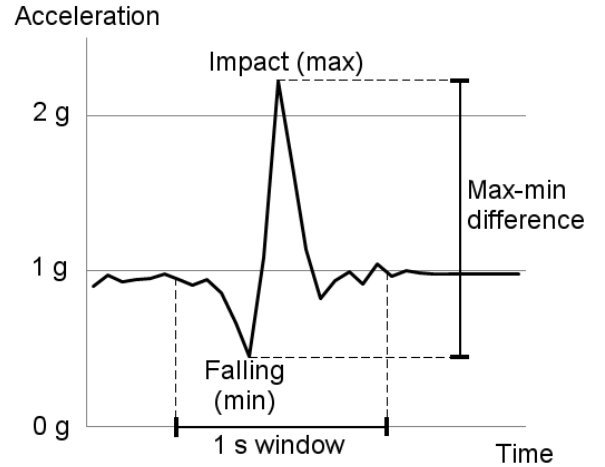


Figure 5. Acceleration pattern during a fall.

III. EVALUATION

A. Controlled Lab Experiment

To evaluate the methods in the RAReFall system, we used a 90-minute activity scenario recorded by 10 people. The scenario contained several sub-scenarios: walking on a treadmill, cycling on a stationary bicycle, elementary activities such as: sitting, lying, standing, and specialized activities such as: cooking, reading, typing, washing dishes, and scrubbing the floor. Eventually, the AR method was evaluated on more than 42,000 activity data samples. The method evaluation was performed with the leave-one-person-out cross-validation technique: models were trained on the data of nine people and tested on the remaining person. This procedure was repeated ten times. Therefore the testing person was completely unknown to the trained model, which is an advantage if the model is intended to be used on previously unknown people.

A summary of the results is shown in Table 1, which shows that the RAReFall system successfully recognized the basic activities with high accuracy: 99% and 98% the Hybrid and the DM model trained with RF (DM-RF), respectively [14]. These results encouraged us to take part in the EvAAL-AR competition.

Table 1. Summary of the recognition accuracy achieved on the controlled lab tests.

	DM			Hybrid
	Decision Tree	Naive Bayes	Random Forest (DM-RF)	
Accuracy	96.36%	97.93%	98.25%	99.04%

B. EvAAL-AR Live Competition

EvAAL-AR is a live competition, which evaluates ARS using 5 criteria: recognition performance, recognition delay, installation complexity, user acceptance and interoperability. Therefore, the RAReFall system was prepared to address each of them. In particular, the recognition performance was already addressed with the controlled lab experiments and the Hybrid method was chosen as the best performing one. The delay criteria was also already addressed, because the method was able to recognize activities with 1 second delay (because of the sliding window). The user acceptance level was addressed by minimizing the number of sensors to two and putting them into pockets on the user’s clothes (instead of using uncomfortable elastic straps to fix them). The installation complexity of the system was minimized because it took less than a minute for the user to put the sensors and start the system on the PC. The interoperability criteria was addressed by first documenting all of the software and making it freely available either under BSD license or direct download of the packed library – <http://dis.ijs.si/chiron/evaal/>. Then, by using communication standards such as Bluetooth and TCP/IP sockets. Finally, GUI application was created, which shows the status of the sensors and the live recognition of the activities. In future, we plan to provide API or even support web-service, so that the system can be connected to other applications.

Table 2. EvAAL-AR overall results.

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

Table 2 shows the scores on the scale of 0–10 for the 2012 and 2013 competition editions. The RAReFall system obtained the highest final score for both years, by achieving not only high accuracy, but also scoring well on the other criteria [13].

The RAReFall system was evaluated as best [13], achieving average score of 83.6%, which was for 14.2 p.p. better than the second-place system, and for 12.9 p.p. better including the results of the previous year competition. It outperformed the other systems in almost each criteria. The biggest problem was the recognition performance, where it achieved 78.7% accuracy (shown in Table 3), which is significantly lower compared to the pre-competition accuracy, i.e., 99%.

Table 3 shows a confusion matrix, the recall for each activity, and the overall accuracy. The results show that walking, standing and lying are correctly recognized most of the time, i.e., 93.1%, 93% and 100% recall values, respectively. Cycling is recognized with 58.9% recall, which is not as good as in the controlled lab experiments [14]. The problem with the cycling was that in the training data it was performed more vigorously and at the competition the subject was cycling slowly. This example shows why DM models can fail in situations different from the ones included in the training data.

Table 3. Confusion matrix – Hybrid model.

	Walking	Standing	Sitting	Lying	Bending	Cycling	Falling	Recall
Walking	322	19	5	0	0	0	0	93.1%
Standing	11	172	2	0	0	0	0	93.0%
Sitting	0	0	35	115	0	0	0	23.3%
Lying	0	0	0	187	0	0	0	100%
Bending	0	0	16	0	0	0	0	0.0%
Cycling	3	0	28	0	8	56	0	58.9%
Falling	1	0	0	0	3	0	6	60.0%
Accuracy								78.7%

Even though falling was recognized with 60% recall, our system successfully recognized the single fall event performed at the competition and no false alarms were raised. The four misclassified instances in Table 3 are the ones that occurred during the falling. The competition’s rules required that the fall event should be recognized as 10 consecutive instances of falling (5 seconds). This was one of the flows which was addressed by the competitors and the competition’s evaluation committee and was suggested that it would be reconsidered in future competitions.

Probably the biggest problem is the poorly recognized sitting activity, i.e., only 23.3% instances were correctly recognized, the other 76.7% were recognized as lying. The reason is that in the controlled lab tests the torso accelerometer was firmly fixed with an elastic strap and at the competition the sensor was put in an elastic pocket at the abdomen, which resulted in slight rotations while performing various activities. In particular, the torso-accelerometer was rotated towards the subject’s body (Figure 6 – C), causing the system to recognize the activity as lying. To confirm that this is not due to wrong adjustments of the expert rules, we performed additional experiments using single DM model trained with RF (DM-RF). The results are shown in Table 4 and they are similar to the ones achieved by the hybrid model, i.e., sitting is recognized as poorly as shown in Table 3. This additionally confirms the fact that the torso accelerometer rotates while performing other activities and the adjustment of the rules is not the problem.

At the competition, the same activity scenario was repeated twice, with the better run counting towards the final score. Summary of these results can be seen in Table 5, which shows that the results for both runs are similar, i.e., sitting is recognized as lying because of the rotations of the torso sensor.

Table 4. Confusion matrix – DM-RF model.

	Walking	Standing	Sitting	Lying	Bending	Cycling	Falling	Recall
Walking	322	11	0	0	9	4	0	93.1%
Standing	13	168	0	0	3	1	0	90.8%
Sitting	0	0	41	109	0	0	0	27.3%
Lying	0	0	0	187	0	0	0	100%
Bending	0	0	16	0	0	0	0	0.0%
Cycling	5	1	7	0	27	55	0	57.9%
Falling	3	0	0	0	1	0	6	60.0%
Accuracy								78.8%

To summarize, Table 3 and Table 4 show how an ARS which recognizes activities with 99% accuracy can significantly drop the accuracy once employed in more realistic conditions, i.e., it is evaluated in a living lab on scenario different from the one used for creation of the AR model.

The dataset, which was collected at the EvAAL competition and is also used for this study is publicly available: <http://dis.ijs.si/ami-repository/index.php?d=14>.

IV. POST-COMPETITION IMPROVEMENTS – CALIBRATION

In order to solve the problem of the rotation of the torso-sensor while performing different everyday activities, we developed a calibration method that corrects accelerometer axes orientation by applying a rotation transformation to the device's raw data. This rotation compensates for user variability in device alignment when standing or walking naturally. The standing position coordinate system for the torso accelerometer is described in Figure 6: (A) normal orientation – front view, (B) normal orientation – side view, (C) rotated orientation – side view.

Figure 6 shows the 3D coordinate system (x, y and z) and the sensed acceleration (in this case the Earth's gravity – g). To calculate the angle between the actual acceleration (e.g. the Earth's gravity (g) for static activities) and some of the axis (e.g., x-axis) we used the previously shown equation (1).

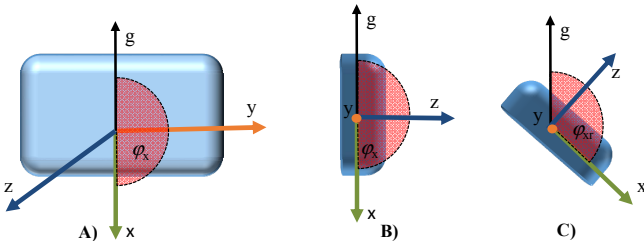


Figure 6. (A) Normal orientation of the sensor while standing – front view; (B) Normal orientation of the sensor while standing/walking – side view, (C) Rotated (misaligned) orientation of the sensor while standing/walking – side view.

As shown in Figure 6, the angle between the g and the x-axis is different (cases B and C) and causes the misclassifications. Using rotation matrices and trigonometry

one can easily rotate the sensor's coordinate system, in such a way that it will correct the data. In order to do this, one should calculate the difference between the expected angle (φ_x) and the rotated angle (φ_{xr}). This difference gives the angle by which the coordinate system should be rotated in order to correct the orientation of the sensor. The rotation is performed by a rotation matrix, which describes a rotation of a coordinate system with respect to another orientation. An acceleration vector in the initial reference frame can be transformed into a vector in a rotated frame by multiplication of the initial vector with the rotation matrix [17]. In three dimensions, rotations are possible around the three principal axes (x, y and z). In this particular case (as shown in Figure 6) the rotation around y axis is needed.

Other calibration techniques were used for preprocessing accelerometer data. Usually they require some intervention from the user [16] or are applied on smartphone sensors data [17]. In our implementation, the calibration is performed using specialized wearable devices, automatically (with no intervention from the user) and the rotation angle ($\varphi_r = \varphi_x - \varphi_{xr}$) is computed in real-time. To achieve this, we used standing or walking as a reference to compute the current angle φ_{xr} . As a reference angle (φ_x) we used the angle computed at the beginning of the activity scenario. The assumption here is that at the start the user fixes the sensor and the sensor has the right orientation. In the EvAAL-AR competition recordings, that means we used the standing sequence (approx. 5 seconds) at the beginning of the scenario. The method then calculates the rotation angle every time a standing/walking activity sequence is recognized. Once it is calculated, all raw accelerometer data thereafter are multiplied by the rotation matrix to achieve the corrected orientation.

Table 5 shows the results achieved by the both methods (DM-RF and Hybrid) for the two runs, on the original data (before applying the calibration method). The results show that the sitting activity is poorly recognized by the both methods for the two runs (from 23% to 29% recall).

Table 5. The recall and accuracy achieved by both methods (DM-RF and Hybrid) for the two runs on the original data (before applying the calibration method).

Original data	First run		Second run		
	DM-RF	Hybrid	DM-RF	Hybrid	
Recall	Walking	93.1%	93.1%	93.9%	93.6%
	Standing	90.8%	93.0%	91.1%	92.6%
	Sitting	27.3%	23.3%	24.3%	28.5%
	Lying	100%	100%	100%	100%
	Bending	0.0%	0.0%	16.7%	0.0%
	Cycling	57.9%	58.9%	29.1%	47.6%
	Falling	60.0%	60.0%	75.0%	75.0%
Accuracy	78.8%	78.7%	73.2%	75.9%	

Table 6 shows the recall and accuracy achieved by both methods (DM-RF and Hybrid) for the two runs, after applying the calibration method. The results show that there is

significant improvement in the overall accuracy in both runs for the Hybrid method and in the second run for the DM-RF method. The detailed results for the Hybrid model for the first run show that the recall value for the sitting increased for 51.4 percentage points (p.p.), from 22.3% to 74.7%. Also the overall accuracy increased for 7.1 p.p., from 78.7% to 85.8%. The improvements are even more emphasized for the second run, i.e., the accuracy increased from 75.9% to 84.3%, and this mainly due to the increased recall for the sitting, i.e., from 28.5% to 80.6%. Additionally, the results are of particular importance because the method addressed exactly the issue with the rotation of the accelerometer during sitting and did not cause significant drop in the recall values for the other activities. Moreover, the recall of the other activities is unchanged, which confirms that the sensor misalignment caused the false recognition of sitting as lying.

Table 6. The recall and accuracy achieved by both methods (DM-RF and Hybrid) for the two runs, after applying the calibration method.

Calibrated data		First run		Second run	
		DM-RF	Hybrid	DM-RF	Hybrid
Recall	Walking	85.5%	91.6%	93.4%	93.6%
	Standing	89.7%	93.0%	91.1%	92.1%
	Sitting	47.3%	74.7%	85.4%	80.6%
	Lying	100%	99%	100%	100%
	Bending	0.0%	0.0%	16.7%	0.0%
	Cycling	57.9%	58.9%	28.2%	47.6%
	Falling	60.0%	60.0%	75.0%	75.0%
Accuracy		79.0%	85.8%	82.8%	84.3%

The results for the DM-RF model are not that convenient as for the Hybrid model. In particular, for the first run with the calibrated data there is no improvement in accuracy compared to the original un-calibrated data, i.e., 78.8% compared to 79%. An improvement is noted for sitting (from 27.3% to 47.3% recall), however this is compensated with the worse performance for walking, which is misclassified as bending. This shows the advantage of using the Hybrid model when the activity can be well described with expert rules (in this case bending).

V. DISCUSSION AND CONCLUSION

The paper presented the general problem with evaluation of ARS, i.e., an ARS may perform perfectly in controlled lab experiments but significantly worse once applied to more realistic conditions: different environment, different activity scenario, etc. One of the reasons for this is that DM models (as one of the most commonly used approaches for AR) perform and generalize well if the training data fully represents the problem domain, i.e., the training data should include all of the user's postures and activities. Practically, this is almost impossible task – to include all possible postures and activities in the training data.

To avoid this issue, we created the Hybrid model for AR. It combines expert rules to recognize postures (static activities),

which ensure the behavior of the system is predictable and robust, and RF classifiers to recognize dynamic activities, for which the rules are not sufficiently accurate. However, even with this improvement, the RAReFall system failed to maintain the accuracy level achieved at the controlled lab experiments. The main problem was in the practical usage of the system, i.e., because of the volunteer's body anatomy and the elastic clothing, the torso accelerometer was slightly rotated towards the body, causing the sitting to be misrecognized as lying. To solve this problem we developed an automatic calibration method that adapts and re-calibrates the accelerometers data in real-time while the user is performing everyday activities. This modification significantly increased the recognition performance, i.e., for 8 p.p. in overall accuracy and for 51 p.p. for the sitting.

Even though the proposed calibration method significantly improved the accuracy for the sitting activity, cycling and bending were still poorly recognized. A confusion matrix of the recognized activities for the Hybrid method for the first run is shown in Table 7. The problem with the cycling was that it was confused with sitting. The reason for this is that in the training data the cycling was performed more vigorously and at the competition the subject was cycling slowly. A possible solution is to apply expert rules also for the cycling activity and not a DM model (as it is in the current implementation of the Hybrid model). The zero recall value for the bending activity was because of the different interpretation of the bending posture. In particular, we described this activity (with expert rules) as a posture where the user is leaning with his legs in upright position. At the competition the volunteer was kneeling and taking some object from the ground. This can be solved by adjusting the expert rules.

Table 7. Confusion matrix – Hybrid model, calibrated data, first run.

	Walking	Standing	Sitting	Lying	Bending	Cycling	Falling	Recall
Walking	317	19	5	0	5	0	0	91.6%
Standing	11	172	2	0	0	0	0	93.0%
Sitting	0	0	112	38	0	0	0	74.7%
Lying	0	0	0	186	1	0	0	99%
Bending	0	0	16	0	0	0	0	0.0%
Cycling	3	0	28	0	8	56	0	58.9%
Falling	1	0	0	0	3	0	6	60.0%
Accuracy								85.8%

Some may argue the importance of distinguishing basic activities such as: sitting, lying and standing. One motivation is that they constitute around 90% of all the activities during a normal day of an average person [21][22]. Furthermore, the American Medical Association (AMA) agrees that sitting for extended periods of time can be bad for personal health: sitting disease, cardiovascular disease, etc. Therefore, having an ARS that recognizes these three activities is essential and may improve a person's health, e.g., by suggesting standing after long period of sitting.

In the future we plan to use the calibration method to benchmark datasets, such as OPPORTUNITY [9] or HASC [10] and additionally confirm the importance of using it. Even though our calibration method significantly improved the recognition accuracy, it is still dealing with only one problem

that can emerge in a real-life situation, i.e., sensor misalignment. There are numerous other issues that need to be tackled in order to create accurate and robust AR system, such as: sensor malfunction, battery issues, dropping the sensor without noticing and similar. These are all practical issues that we plan to tackle in the future, once we deploy the system in more realistic, real-life usage.

ACKNOWLEDGMENT

This work was partly supported by the Slovenian Research Agency under the Research Programme P2-0209 Artificial Intelligence and Intelligent Systems and partly by the CHIRON project - ARTEMIS Joint Undertaking, under grant agreement No. 2009-1-100228. The authors would like to thank Simon Kozina for the help provided while programming the initial activity recognition algorithm.

REFERENCES

- [1] K. Ducatel, M. Bogdanowicz, F. Scapolo, J. Leijten, J. Burgelman. Scenarios for ambient intelligence in 2010.
- [2] H. Gjoreski, M. Gams, M. Luštrek. Context-Based Fall Detection and Activity Recognition Using Inertial and Location Sensors. JAISE, 2014, 6(4), pp. 419-433.
- [3] H.H. Manap, N. M. Tahir, A. I. M. Yassin. Anomalous Gait Detection based on Support Vector Machine. In Proc. International Conference on Computer Applications and Industrial Electronics, 2011, pp. 623-326.
- [4] H. Gjoreski, B. Kaluža, M. Gams, Radoje Milić, M. Luštrek. Context-based Ensemble Method for Human Energy Expenditure Estimation. Applied Soft Computing. In press, 2015.
- [5] M. Altini, J. Penders, R. Vullers, O. Amft, Estimating energy expenditure using body-worn accelerometers: a comparison of methods, sensors number and positioning, IEEE J. Biomed. Health Inf. 99 2014, pp. 219–226.
- [6] M. Gjoreski, H. Gjoreski, M. Luštrek and M. Gams. Automatic detection of perceived stress in campus students using smartphones. Intelligent Environments (IE), (2015), pp. 132 – 135.
- [7] H. Alemdar, C. Tunca, C. Ersoy. Daily life behaviour monitoring for health assessment using machine learning: bridging the gap between domains. Journal of Personal and Ubiquitous Computing, 19 (2): pp. 303-315.
- [8] B.G. Steele, B. Belza, J. Hunziker, L. Holt, M. Legro, J. Coppersmith, D. Buchner, S. Lakshminaryan. Monitoring daily activity during pulmonary rehabilitation using a triaxial accelerometer. J Cardiopulm Rehabil. 2003, 23(2): pp.139-42.
- [9] R. Chavarriaga, H. Sagha., A. Calatroni, S. Digumarti, G. Tröster, J. Millán, D. Roggen. The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition, Pattern Recognition Letters, 2013.
- [10] N. Kawaguchi, N. Ogawa, Y. Iwasaki, K. Kaji, T. Terada, K. Murao, S. Inoue, S. Kawahara, Y. Sumi, N. Nishio. HASC Challenge: Gathering Large Scale Human Activity Corpus for the Real-World Activity Understandings, ACM AH, (2011), pp. 1-5.
- [11] B. Kaluža, S. Kozina, M. Luštrek. The activity recognition repository: towards competitive benchmarking in ambient intelligence. Activity Context Representation, AAAI, (2012).
- [12] EvAAL - Evaluating AAL systems through competitive benchmarking. <http://evaal.aalooa.org/>
- [13] H. Gjoreski, S. Kozina, M. Gams, M. Luštrek, J.A. Álvarez-García, J.H. Hong, J. Ramos, A.K. Dey, M. Bocca, N. Patwari. Competitive Live Evaluation of Activity-recognition Systems. IEEE Pervasive Computing, 14(1), (2015), pp. 70 – 77.
- [14] S. Kozina, H. Gjoreski, M. Gams, M. Luštrek. Efficient Activity Recognition and Fall Detection Using Accelerometers. Evaluating AAL Systems Through Competitive Benchmarking Communications in Computer and Information Science, (2013), pp. 13-23
- [15] S. Kozina, H. Gjoreski, M. Gams, M. Luštrek. Three-Layer Activity Recognition Combining Domain Knowledge and Meta-Classification. JMBE (2013), 33(4), pp. 406-414.
- [16] A. Friedman, N. Hajj Chehade, C. Chien, G. Pottie. "Estimation of accelerometer orientation for activity recognition," Engineering in Medicine and Biology Society (EMBC), pp. 2076-2079, 2012.
- [17] M.D. Tundo, E. Lemaire, N. Baddour. Correcting Smartphone orientation for accelerometer-based analysis, Medical Measurements and Applications Proceedings (MeMeA), (2013) pp. 58-62.
- [18] The Confidence Project, European FP7, 2014. Retrieved September, 2014 from: <http://www.confidence-eu.org>
- [19] The CHIRON project JU ARTEMIS, 2014. Retrieved September, 2014 from: <http://www.chiron-project.eu/>
- [20] H. Gjoreski, S. Kozina, M. Gams, M. Luštrek. RAREFall – Real-time Activity Recognition and Fall Detection System. IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM), live DEMO. 2014; pp. 145–147.
- [21] Just stand organization. <http://www.juststand.org>
- [22] A.G. Bonomi. Physical activity recognition using a wearable accelerometer. PhD Thesis, 2010.