

# Robust Activity Recognition Combining Anomaly Detection and Classifier Retraining

Hesam Sagha<sup>1</sup>, Alberto Calatroni<sup>2</sup>, José del R. Millán<sup>1</sup>, Daniel Roggen<sup>2</sup>, Gerhard Tröster<sup>2</sup>, Ricardo Chavarriaga<sup>1</sup>

<sup>1</sup>CNBI, EPF Lausanne, Switzerland

<sup>2</sup>Wearable Computing Laboratory, ETH Zurich, Switzerland

hesam.sagha@epfl.ch

**Abstract**—Activity recognition systems based on body-worn motion sensors suffer from a decrease in performance during the deployment and run-time phases, because of probable changes in the sensors (e.g. displacement or rotation), which is the case in many real-life scenarios (e.g. mobile phone in a pocket). Existing approaches to achieve robustness tend to sacrifice information (e.g. by rotation-invariant features) or reduce the weight of the anomalous sensors at the classifier fusion stage (adaptive fusion), ignoring data which might still be perfectly meaningful, although different from the training data. We propose to use adaptation to rebuild the classifier models of the sensors which have changed position by a two-step approach: in the first step, we run an anomaly detection algorithm to automatically detect which sensors are delivering unexpected data; subsequently, we trigger a system self-training process, so that the remaining classifiers retrain the “anomalous” sensors. We show the benefit of this approach in a real activity recognition dataset comprising data from 8 sensors to recognize locomotion. The approach achieves similar accuracy compared to the upper baseline, obtained by retraining the anomalous classifiers on the new data.

**Keywords**—*anomaly detection; classifier adaptation; activity recognition;*

## I. INTRODUCTION

Typically, static human activity recognition systems are based on one or more sensor systems where signal processing and pattern recognition methods are applied to recognize activities. This process is applied in different fields, e.g., for assisted living [1] and industrial applications [2]. Body-worn sensors for activity recognition comprise different modalities, particularly acceleration sensors are extensively used [3]. The recognition system is usually trained beforehand by collecting datasets comprising the activities of interest, with a fixed sensor setup. The usual assumption behind any pattern recognition approach is that training data are as similar as possible to the final data which the system will have to classify once it is deployed, which is a very restrictive assumption [4]. For instance, in the case of body-worn sensors, this would mean that sensors need to be placed and oriented exactly in the same positions as in the training phase, which becomes very impractical for the user in daily life. To understand how critical this issue is, think about performing activity recognition with a smartphone [5], whose orientation in the pocket is different from time to time. The acceptance by users would indeed be reduced, if they had to take care of where and how they place their phones.

If sensors are anomalous (they are not behaving as they are expected), the classifier models do not match anymore

the data measured by the sensors and there is a performance degradation, whose amount will depend on the entity of the imposed anomaly<sup>1</sup>. Most of the approaches proposed so far to address this issue try to introduce countermeasures to limit the degradation, and minor study is done on the adaptation of classifiers, or to adapt the system to a new equilibrium point, where the system converges to a new status where the anomalous sensors are considered as “normal”. In this paper we propose to follow the later approach as it would bring some advantages:

- The system is then able to continue using the data delivered by the anomalous sensors, if they are still useful for the activity recognition task, which should allow for higher accuracy;
- After adaptation, the system is again ready to bare with other anomalies in sensor.

Although, the type of anomaly could be different for any sensor modality, in this paper we focus our study on rotation of IMU sensors. However, the methods is independent of the type of anomaly and could be extended to other type of anomalies.

Section II illustrates related work and its limitations. The proposed method is described in detail in Section III. Section IV presents the dataset and the experiments used for validation and illustrates the comparison to the previous anomaly detection and rectification approach. The results are then presented in Section V and discussed in Section VI. Conclusion and outlook follow in Section VII.

## II. RELATED WORK

In the literature, the problem of anomalous sensors in activity recognition has been tackled following three main strategies:

- Extraction of robust features;
- Collection of data in different conditions;
- Adaptation of the signal processing chain.

For example, in the case of sensor displacement or rotation, the first approach, followed by [6], has the disadvantage of losing the information on the orientation of the sensors, which can be necessary for recognizing fine-grained activities or

---

<sup>1</sup>The term “Anomaly” is more general than “noise” or “outlier”. It is looking into the trend or behavior of the data. E.g. an IMU sensor could be noiseless, but is experiencing a rotation.

gestures. Extracting robust features with respect to certain type of anomaly (sensor displacement) is studied in [7], with extra burden on feature extraction phase and may not be generalized to other type of anomalies.

The second approach [8] can become quite cumbersome if multiple positions and rotations and combinations thereof need to be included. Furthermore, the accuracy obtained is lower compared to having a single specialized model for the correct position and orientation and the degradation could increase by adding more and more configurations.

The third approach can be carried out by adapting classifiers to the data trend in an unsupervised manner [9] which could make the classifiers unstable, even when there is no noise. Another way is to first detect which sensors are anomalous and take a counteraction. Detection of anomalous classifiers<sup>2</sup> are proposed at fusion level by monitoring the consistency between outputs of the classifiers operating on each sensor system [10]. This work proposes the use of classifier fusion architecture where each classifier is responsible for a subset of sensors. When the decisions provided by one or more classifiers are not anymore in line with the joint distribution of the decisions learned in a training phase, an anomaly is detected and the responsible classifiers are flagged. Once classifiers are flagged as anomalous, different actions can be taken:

- It has been shown that the automatic removal of the anomalous classifiers from the fusion mitigates the loss in performance in terms of recognition accuracy [10]. This approach is limited due to the following fact: if the classifiers are monitored over a long time span and are then removed one by one, at the end not enough classifiers are left to get a good classification performance, while the sensors are still available and may be capable of transmitting reasonable data or features (eventually just rotated).
- Adaptation of fusion parameters is another way to cope with the anomalous classifiers. An incremental approach has been proposed in [11] for different trainable fusion types (Fuzzy Integral, Decision Template, Dempster-Shafer). This approach may not be able to handle high levels of noise, e.g., when an anomalous classifier is completely biased toward one class. In such case, fusion adaptation may give at most a similar performance to what obtained through removal, because the classifier does not add any information to the fusion.

These methods do not make use of the fact the anomalous sensors are more often than not still delivering useful data (e.g. a sensor is rotated), which with the previous approaches are either not used anymore, or are having less weight in the fusion. Furthermore, fusion level anomaly detection allows to exploit all available sensors even if they have changed at runtime.

---

<sup>2</sup>The terminology 'anomalous classifier' is used to separate it from 'anomalous sensors'. An anomalous classifier dictates the corresponding sensor is anomalous. But, not necessary an anomalous sensor results abnormal behavior in the classifier outputs, e.g. small rotation.

- We go beyond the state of the art by introducing a method which combines the ability of the previous approaches to detect the anomalous sensors and adapts the classifier models. This not only attempts to minimize the performance loss, but also brings the system to a new stable state. This is accomplished through a self-training process, which uses the fusion output to provide labels to retrain the sensor systems flagged as anomalous. With this approach, the system can then again react to new anomalies and maintains resilience in case sensors are subsequently removed from the fusion (e.g. systems switched off).

### III. METHOD

We propose to cascade anomaly detection (AD) capabilities and retraining of the recognition chain. The AD is based on an information-theoretic approach which has been fully described in [10]. The AD models the behavior of the classifiers by computing the mutual information between their decisions. These values are monitored during runtime and their difference w.r.t. the ones evaluated in the training set are computed online. When the difference exceeds a predefined threshold, the corresponding classifiers are flagged as anomalous. Runtime values are computed in a moving window which keeps  $w$  previous decisions.

In order to be robust against fast temporary changes and false positives, the anomaly detection does not immediately lead to a counteraction, but a measure of the sensor reliability (Trust Indicator: TI) is computed continuously for each classifier based on the AD output. At each sample, the TI value is decreased/increased for all classifiers which have been flagged as anomalous/healthy. Only when the TI value falls under a certain threshold  $t$ , the system triggers an action: adaptation of the classifier parameters and/or retraining.

Retraining the recognition chain consists of retraining both the anomalous classifiers and the fusion block, Fig. 1. The anomalous classifiers are adapted to the new data trend coming from the sensors. To do this, the system needs class labels, which are not available at run-time. An approximation of the class labels is obtained via the fusion decision. The fusion decision is made by ignoring the anomalous classifiers, which yields better accuracy compared to fusing all classifiers [10]. Upon detection of the anomalous classifiers, the system forgets their parameters and it incrementally recalculates them with the newly chunks of acquired activity examples [12]. For example in the case of Nearest Centroid Classifier (NCC) which we used, the center of the detected class is moved slightly toward the new data sample.

There are two types of classifier fusion: non-trainable (e.g. Borda count, majority voting, sum-rule), or trainable (e.g. Bayesian fusion, Decision Template, Fuzzy Integral, Dempster-Shafer) which have parameters to be estimated [13]. In the former case, only the classifiers are adapted to the new data. While, in the later case, the fusion parameters should be adapted. Since often fusion parameters of the various classifiers are independent from each other, the same approach (forgetting and recalculating) could be applied to update the fusion parameters [11]. In this case, to avoid biasing the recognition chain, the classifiers and fusion block are trained

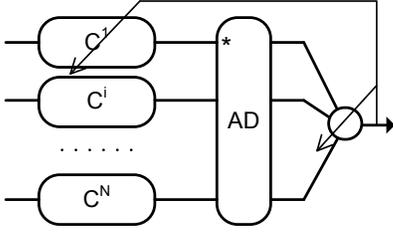


Fig. 1. Adaptation process. (\*) indicates detected anomalous classifier. Fusion block should be updated in the case of trainable fusion methods.

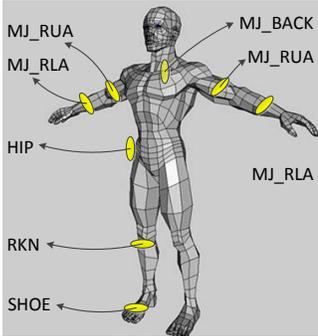


Fig. 2. Sensor setup. All depicted sensors are triaxial accelerometers. Five of them are mounted within a “motion jacket” (prefix “MJ”) for a reproducible deployment.

with different activity examples: this could be achieved by interleaving chunks of data, one part for adapting classifiers and the other for adapting fusion parameters.

In this paper we focus on classifier retraining without adapting fusion block.

#### IV. EXPERIMENT

We test the proposed approaches on the data of three subjects from the OPPORTUNITY dataset [14]. This dataset contains gestures and activities performed in a kitchen multiple times (in so-called “runs”) and recorded with body-worn and ambient sensors. We selected five runs of each subject (tagged ADL1-ADL5) and each run is approximately 20 minutes. The data are publicly available and were also used for a challenge at the SMC 2012 conference [15]. Out of the pool of sensors available in the dataset, we used eight triaxial accelerometers (the sensor setup and abbreviations are shown in Fig. 2). Five of them were mounted within a “motion jacket” at the lower and upper arms and on the back; the remaining three accelerometers were placed on one shoe, on one knee and on a hip of the subject.

Of the five runs within the dataset, the first two are used for training the classifiers and the anomaly detection parameters. Without loss of generality, we tested only the method on rotational noise, since the anomaly detection performance on other noise type (additive noise) is provided in [10]. Runs ADL3 to ADL5 are artificially corrupted by rotational noise. The amount of noise is 10, 40 and 60 degrees in a random direction on 3 axes of the accelerometers and the results are based on the averaged over 10 noise realization and three subjects. The affected sensors are provided in table I for different

TABLE I. NOISY SENSORS FOR DIFFERENT SIMULATIONS.

Sim	Noisy sensor(s)
1	Right Knee (RKN), MJ Back
2	Right Knee, MJ Back, MJ RUA
3	Right Knee, MJ LLA, Right Shoe, Hip
4	Right Knee, MJ Back, MJ RUA, MJ LUA
5	Right Knee, MJ Back, MJ LLA, Right Shoe

TABLE II. USE OF EACH RUN.

Run	Usage	Noisy
ADL1	Training	
ADL2	Training	
ADL3	Adaptation	+
ADL4	Adaptation	+
ADL5	Testing	+

runs. Out of the whole set, these sensors are selected so that the performance degradation is more sensible. The anomaly detection procedure continuously monitors the decisions within the runs ADL3-4 and the adaptation and/or retraining start (if necessary) to adapt the system to the new condition. The performance during the whole adaptation process is evaluated at each time step on ADL5 with the models obtained in the adaptation process. Table II shows a summary of the data usage.

The class labels correspond to the modes of locomotion (walk, sit, stand, and lie). The Nearest Centroid Classifier (NCC) is used for each 3-axis accelerometer, and the decisions of all classifiers are fused with a majority fusion approach. The features for each stream are the mean of 0.5 second of data. We expect a performance loss at the beginning of the adaptation session, and a gradual increase in the performance after the anomaly detection block detected the faulty classifiers. We present the evolution of the accuracy during adaptation for each single classifier, as well as for the fused decision.

We compare the newly proposed approach with the following baselines:

- Lower bound (LB): a worst-case approach, which consists in not performing any adaptation, meaning that the classifiers operating on the rotated sensors will deliver mostly wrong labels (this would be how a non-adaptive system would operate in reality after some sensors get rotated or displaced).
- Upper bound (UB): consisting in retraining classifiers associated with the rotated sensors manually, using class labels, so that the models are perfectly adapted to the new condition. This is the upper limit that the system should achieve in case the adaptation is completely successful.

#### V. RESULTS

The results for one subject are presented in Fig. 3. The noisy sensors which are rotated by 40 degrees are Right\_Knee, MJ\_Back, MJ\_RUA, MJ\_LUA (4th Sim). Noisy Sensors detected as anomalous are shown as black strip in Fig. 3(a). It is once a TI value falls beyond the threshold of 0.3 and at this point retraining starts. It can be seen that two of the noisy classifiers are pretty well detected as anomalous after 430 samples (13 min) and the other ones around sample 650 (19 min) and 1200 samples (35 min). Note that window length of

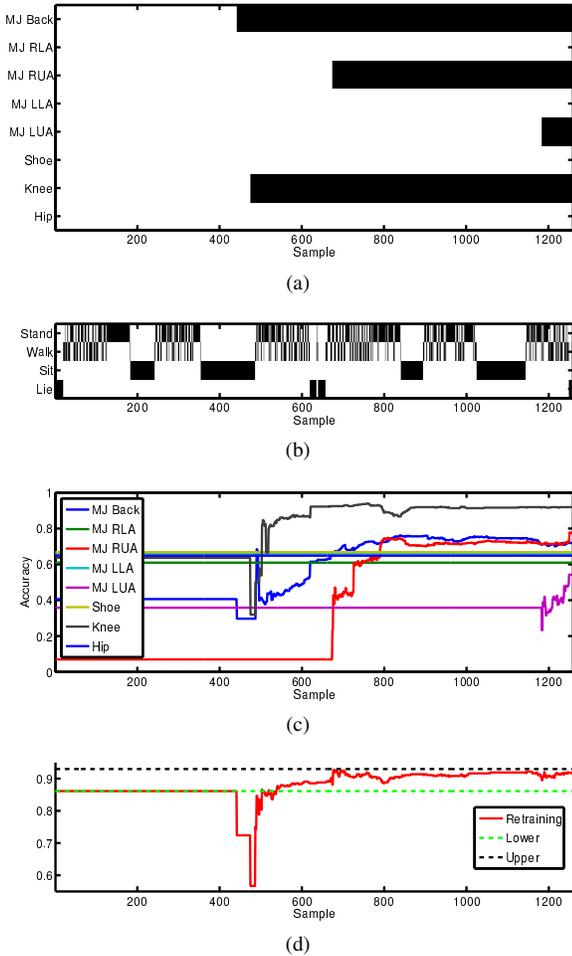


Fig. 3. Detection and adaptation for Sim 4 with noise of 40 degrees. a) Detection of anomalous classifiers. A classifier is detected as anomalous once its TI value passed the threshold (0.3). b) Class labels. c) Evolution of the accuracy of each classifier, at each time the accuracy is computed on the test set with the current updated parameters. d) Accuracy of the fusion on test set while retraining classifiers. 200 Samples  $\approx$  6 minutes.

AD,  $w = 400$  (12 min), and before that no detection is carried out.

The accuracy progression of the classifiers during the adaptation phase is depicted in Fig. 3(c). The performance computed for each point is on the whole test set. The performance of the anomalous classifiers gradually increases after each sample arrives. Once a sample from a new class is introduced to the classifiers, there is a jump in the performance.

Figure 3(d) represents the fusion performance on the test set after classifier retraining. At first, there is a huge performance degradation due to that the anomalous classifiers have not yet received enough data samples but still they are in the fusion process.

#### A. Anomaly detection accuracy

The performance of the anomaly detection algorithm is shown in table III for locomotion recognition. TP is the rate of correctly detected rotated sensors, and TN is the rate of

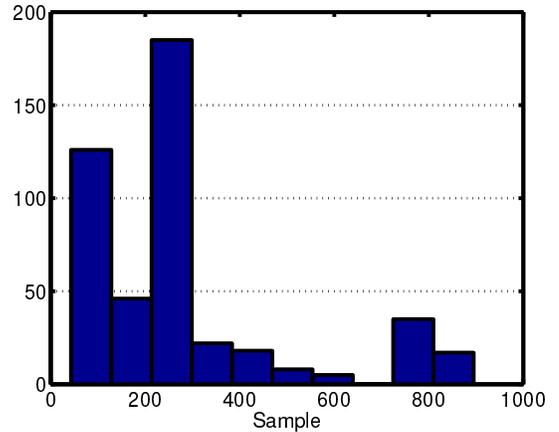


Fig. 4. Histogram of the delay of anomaly detection.

TABLE III. ANOMALY DETECTION ACCURACY.

Sim	Rotation					
	10°		40°		60°	
	TP	TN	TP	TN	TP	TN
1	0.07	1	0.23	1	0.37	1
2	0.10	1	0.40	1	0.45	1
3	0	1	0.24	1	0.23	1
4	0.07	0.99	0.26	1	0.32	1
5	0.07	1	0.22	1	0.22	1

not-detecting unchanged sensors, ignoring the first  $w$  samples. It can be seen that the anomaly detection performance is correlated with the level of noise. This is not a major limitation, since a low level of noise will also affect the decisions of the single classifiers and of the fusion block only marginally. On the other hand, AD has a very high TN, irrespective of the level of noise. A comparison shows that the sensors which are more involved in the classification (Right\_Knee, Back, RUA) are being detected better. Increasing the number of noisy sensors makes detection process more difficult, because less healthy sensors are available for comparison. On the other hand, higher rotation level makes the detection easier, because faulty classifiers are less in accordance with the healthy ones. The histogram of the delay of detection is provided in Fig. 4. Latency in detection is one main causes of low TP value. Often anomalies are detected within less than 300 samples (9 min), and it is dependent on the chosen parameters such as threshold over TI value.

#### B. Retraining performance

Once the sensors are anomalous the fusion performance once the classifiers are trained on training and on adaptation data (noise included) is provided in table IV. We removed simulations where upper-bound is lower than lower-bound, meaning training based on the new data will decrease the performance. One reason is that there is not enough representative data to have a good classifier.

After anomaly detection and retraining classifiers, the rates of achievement to the upper bounds are shown in table V. It is defined as:

$$(F - LB)/(UB - LB) \quad (1)$$

where  $F$  is the weighted  $F_1$ -measure averaged over the number of samples per class, the same way as LB and UB are

TABLE IV. AVERAGE ACCURACY DIFFERENCE BETWEEN UPPER BOUND AND LOWER BOUND (%). THE PERFORMANCES ARE OBTAINED ON THE TEST DATASET.

Sim	Rotation		
	10°	40°	60°
1	2.85	5.06	6.00
2	3.66	6.35	11.60
3	2.91	9.04	14.24
4	4.03	15.53	24.04
5	2.14	9.45	14.28

TABLE V. RETRAINING: RATE OF ACHIEVEMENT TO THE UPPER BOUND.

Sim	Rotation		
	10°	40°	60°
1	0.04	0.09	-0.11
2	0.10	0.13	0.62
3	0	0.24	0.45
4	0.10	0.53	0.62
5	-0.08	0.80	0.57

calculated. For high level noises the rate of achievement is quite considerable, except for Sim 1, which we found it is due to the latency of AD, and it couldn't adapt with the remaining samples, which is also the reason of negative rates.

## VI. DISCUSSION

The proposed approach has advantages and limitations; we here discuss the main points:

- Stopping criterion: an important question is when to stop adaptation. The most straightforward solution is to just stop the retraining phase after a certain number of samples for each activity class have been recognized. The number of samples needed to achieve reasonable classifier models is application-dependent and depends also on the used classifier.
- Reachable accuracy: The retraining is not guaranteed to yield good classifier models, since it gets the labels from the fusion, which is not perfect. That is, if the fraction of correct labels provided by the fusion block is not high enough, then the models built on the rotated sensors will not be good.
- Computational cost: retraining the classifiers can become computationally expensive in case of algorithms like support vector machines, where the training phase for  $N$  examples has a complexity between  $O(N^2)$  and  $O(N^3)$  (depending on the implementation). Alternative solutions involving just the adaptation of the fusion block would be attractive, since they would spare this retraining. Furthermore, retraining is executed in a distributed fashion on different sensors, involving battery usage, whereas the fusion is usually performed in a central "super-node", which might have a bigger battery or even a supply directly from a power outlet.
- Risk of failures: under the hypothesis of a rotation (or any other anomaly which does not prevent the sensors from delivering data which are still usable for the recognition task) the classifier retraining allows to adapt the system behavior to maintain the performance close to the upper achievable baseline. Nevertheless, if the source of the anomaly is a hardware failure, or another form of noise, retraining the classifier models

on the newly available data might just create a model which does not work. A mechanism operating directly at the signal level would also be needed to estimate whether the data are still usable or not. This can be done by using a metric for evaluating how good the feature space of the anomalous sensor is for the recognition task (e.g. by monitoring the Fisher Score).

- Long-term operation versus short-term anomalies: in a realistic scenario, the system should be able to cope to several subsequent anomalies, e.g. the users rotate many times the mobile phone in their pockets, or different sensors are affected one after the other. Retraining the classifiers, along with adapting the fusion and the anomaly detection parameters, should allow for long-term operation, since this strives to guarantee a certain continuity of service. In fact, after each anomaly, the system would adapt completely, so that after some time it would be ready to cope with the next anomaly. On the other hand, if an anomaly is temporary (e.g. a short rotation of a sensor) it may make sense to just temporarily lower the weight of that sensor in the fusion, instead of retraining the related classifier. Finding the best solution for deciding when to trigger retraining is an open challenge: we believe that it is reasonable to adapt the fusion parameters to cope with the anomaly and only when a classifier is deviating too strongly from the expected behavior its retraining could be triggered. A deviation from the expected behavior can be detected by using a confusion matrix calculated offline and comparing it with one calculated online by treating the fusion output as a ground truth.
- System delay: in order to detect an anomaly, the system needs to first examine many decisions performed by the classifiers and fusion. If the activities of interest are taking place sporadically, this could potentially last a long time and eventually, the adaptation could start too late. To mitigate this, detecting the anomalies at the signal level could be beneficial, and make it faster, since the classifier decisions are obtained from a segment of data. E.g. by monitoring relationships between the signals delivered by sensors. This can work if sensors are close to each other and they deliver data of related modalities (e.g. acceleration and position).

## VII. CONCLUSION

In an open-ended pattern classification task, such as human activity recognition performed with wearable sensors, unpredictable sources of noise and failures can impair the recognition performance. In order to deliver a certain continuity of service, it is inevitable to deploy adaptive methods. We have devised approaches to cope with anomalies which can occur at run-time. An anomaly detection process monitors the decisions from different sources (sensor systems which classify features) and their fusion and informs the system when one or more classifiers are not behaving as expected. Upon detection of these anomalies, we proposed to perform a retraining of the classifiers and fusion parameters. The classifiers and fusion are retrained from scratch with the new trend of "anomalous"

data, while labels obtained at the fusion output are used as ground truth. We showed on a subset of the OPPORTUNITY dataset that retraining can achieve results at the level of the upper baseline, which would correspond to manually training the system with the sensors already in the anomalous state (e.g. rotated).

In future work we will investigate methods operating at the signal level in order to be able to detect anomalies more promptly and eventually also to classify the kind of anomaly that affects the sensors, in order to decide if a retraining is useful at all. Currently, the parameters of the anomaly detection algorithm itself are not adapted: this could be done after waiting that a certain number of samples of all the classes are gathered in order to have a precise estimation.

#### ACKNOWLEDGMENT

The project OPPORTUNITY acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 225938.

#### REFERENCES

- [1] R. Chavarriaga, H. Sagha, H. Bayati, J. d. R. Millán, D. Roggen, K. Förster, A. Calatroni, G. Tröster, P. Lukowicz, D. Bannach, M. Kurz, G. Hölzl, and A. Ferscha, "Robust activity recognition for assistive technologies: Benchmarking machine learning techniques," in *Workshop on Machine Learning for Assistive Technologies at the Twenty-fourth Annual Conference on Neural Information Processing Systems (NIPS)*, 2010.
- [2] T. Stiefmeier, D. Roggen, G. Ogris, P. Lukowicz, and G. Tröster, "Wearable activity tracking in car manufacturing," *IEEE Pervasive Computing Magazine*, vol. 7, no. 2, pp. 42–50, 2008.
- [3] K. Van Laerhoven, A. Schmidt, and H. Gellersen, "Multi-sensor context aware clothing," in *6th International Symposium on Wearable Computers (ISWC 2002)*, IEEE Computer Society. IEEE Computer Society, 2002, pp. 49–56.
- [4] D. Roggen, P. Lukowicz, A. Ferscha, J. d. R. Millán, G. Tröster, and R. Chavarriaga, "Opportunistic human activity and context recognition," *Computer -IEEE Computer Society-*, 2012.
- [5] W. Khan, Y. Xiang, M. Aalsalem, and Q. Arshad, "Mobile phone sensing systems: A survey," *Communications Surveys Tutorials, IEEE*, vol. PP, no. 99, pp. 1–26, 2012.
- [6] K. Kunze and P. Lukowicz, "Dealing with sensor displacement in motion-based onbody activity recognition systems," in *Proceedings of the 10th international conference on Ubiquitous computing*, 2008, pp. 20–29.
- [7] K. Förster, P. Brem, D. Roggen, and G. Troster, "Evolving discriminative features robust to sensor displacement for activity recognition in body area sensor networks," in *5th International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, dec 2009, pp. 43–48.
- [8] J. Lester, T. Choudhury, and G. Borriello, "A practical approach to recognizing physical activities," in *Pervasive Computing*, ser. Lecture Notes in Computer Science, K. Fishkin, B. Schiele, P. Nixon, and A. Quigley, Eds. Springer Berlin / Heidelberg, 2006, vol. 3968, pp. 1–16.
- [9] R. Chavarriaga, H. Bayati, and J. Millán, "Unsupervised adaptation for acceleration-based activity recognition: robustness to sensor displacement and rotation," *Personal and Ubiquitous Computing*, pp. 1–12, To appear.
- [10] H. Sagha, H. Bayati, J. d. R. Millán, and R. Chavarriaga, "Online anomaly detection and resilience in classifier ensembles," *Pattern Recognition Letters*, 2013.
- [11] D. Sannen, E. Lughofer, and H. Van Brussel, "Towards incremental classifier fusion," *Intelligent Data Analysis*, vol. 14, no. 1, pp. 3–30, 2010.
- [12] K. Förster, S. Monteleone, A. Calatroni, D. Roggen, and G. Tröster, "Incremental knn classier exploiting correct-error teacher for activity recognition," in *accepted for The Ninth International Conference on Machine Learning and Applications (ICMLA)*, 2010.
- [13] R. P. W. Duin, "The combining classifier: To train or not to train?" in *ICPR (2)*, 2002, pp. 765–770.
- [14] D. Roggen, A. Calatroni, M. Rossi, T. Holleczeck, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, , G. Pirkl, A. Ferscha, J. Doppler, C. Holzmann, M. Kurz, G. Holl, R. Chavarriaga, H. Sagha, H. Bayati, M. Creatura, and J. D. R. Millán, "Collecting complex activity data sets in highly rich networked sensor environments," in *Seventh International Conference on Networked Sensing Systems*, 2010.
- [15] R. Chavarriaga, H. Sagha, A. Calatroni, S. Digumarti, G. Tröster, J. d. R. Millán, and D. Roggen, "The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition," *Pattern Recognition Letters*, 2013.