On the issue of variability in labels and sensor configurations in activity recognition systems

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

***Abstract*—Two aspects of the design and characterization of activity recognition systems are rarely elaborated in the literature. First, the influence of system performance with variability in sensor placement and orientation is often overlooked. This is important for the deployment of robust activity recognition systems. Second, the influence of labeling variability is also overlooked, especially w.r.t. label boundary jitter and labeling errors. This is important during the development of an activity recognition system as acquiring labels is costly. We argue that there is a need to explicitly address the consequences of such variability in publications, together with the mitigation strategies that are used. Elaborating on this is required to move the state of the art towards real-world applications, such as in industrial wearable assistance applications or pervasive healthcare.**

I. Problem statement

We discuss two aspects towards the design of more robust activity recognition systems for real-world deployment. First, there is a need to clarify system performance under variability of sensor characteristics, in particular placement and orientation variability. This is important as it is difficult, impractical or uncomfortable to ensure precise on-body sensor placement. Second, there is a need to clarify the effect of variability in labeling (especially label boundary jitter and label accuracy). Annotating sensor data is time and cost intensive. We believe that for eventually complex activity or gesture recognition to be deployed in real-world applications there there is a need for strategies and methods to become immune to this variability.

A. Sensor placement and orientation variability

Sensor placement and orientation is usually critical to activity recognition performance. Changes in placement and orientation (hereafter *configuration*) affect the sensor signal patterns corresponding to activities. As classifiers are trained using the sensor signals acquired in a specific configuration, it is important to have the same sensor configuration during operation. This is a constraint that should be relieved:

- It is unrealistic to expect from end-users (e.g. an elderly person) to carefully place and align sensors.
- It is often not comfortable to have always the same sensor configuration. For instance, a user may want to displace a sensor-enabled bracelet, just to let the skin breath.
- Some form-factors should allow for variability, such as sensorized clothing or sensorized ornaments. Users expect sensorized systems to behave as usual (e.g. it should be possible move a sensorized bracelet just as a normal one).
- In the general case, activities should be equally well recognized regardless of where a sensor-enabled device is located, such as on the arm, or in a pocket - so-called "opportunistic sensor configurations" [1].

	$t = s$		$\lvert t - s\rvert = 1$		$\lvert t - s\rvert > 1$	
HCI	84.9%	2.1%	50.0%	21.0%	48.7%	24.4%
Fitness	83.0%	5.7%	65.7%	4.1%	42.0%	9.1%

TABLE I: Classification accuracy when training the classifiers on sensor s and testing them on sensor t which is: the same as s; it's immediate neighbor, or a neighbor further apart. The recognition accuracy quickly drops with larger distance between the train sensor s and test sensor t.

Thus, in real-world applications sensor configuration variability is expected or desired. Current literature often does not explain how a system behaves when the sensor configuration changes. This must be documented in order to:

- Assess the robustness to real-world use of a system
- Compare methods that filter out, or attempt to render the activity recognition chain robust to these variations.

Two examples illustrate the effect of these variations on activity recognition performance. In [2], we investigated the effect of sensor displacement on body segments in two scenarios: HCI gestures (5 classes), and fitness activity recognition (6 classes). For the latter, we recorded data from sensors placed on the lower leg and thigh at regular intervals of about 5cm (see figure 1). Figure 2 illustrates the effect of sensor displacement by showing the activity classes in a two-dimensional feature space. Even for a displacement of only 5cm (two sensors just next to each other), there is already a significant change in the structure of the feature space. Table I shows the that the recognition accuracy drops rapidly as the sensor used during operation is is further away from the sensor used during training.

In [3] we investigated the effect of rotational variability on the recognition performance of manipulative gesture recognition in a car maintenance scenario. During testing, we simulated this variability by rotating the coordinate system of the accelerometers. We investigated the effect of noise with various sensors contributing to activity recognition (fusion of the decisions of discrete HMM classifiers operating on individual sensor nodes). In figure 3 we show increased rotation the activity recognition performance quickly drops. Larger sensor sets partly allows to reduce this effect.

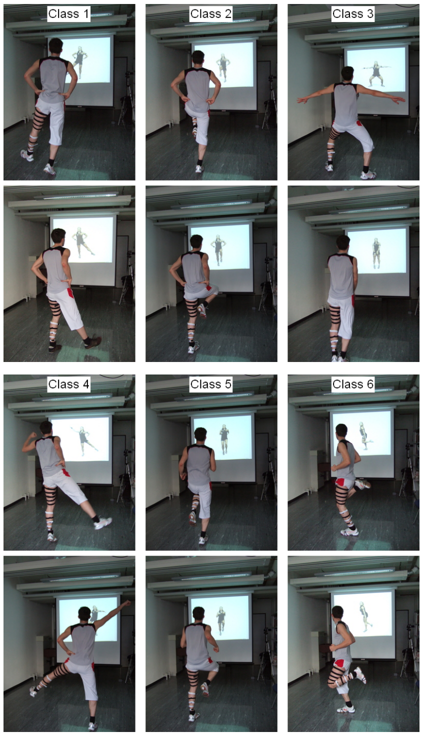

Fig. 1: The fitness scenario includes 6 classes: (1) flick kicks, (2) knee lifts, (3) jumping jacks, (4) superman jumps, (5) high knee runs, (6) feet back runs. For each class, the extent of the body movements is shown on two rows.

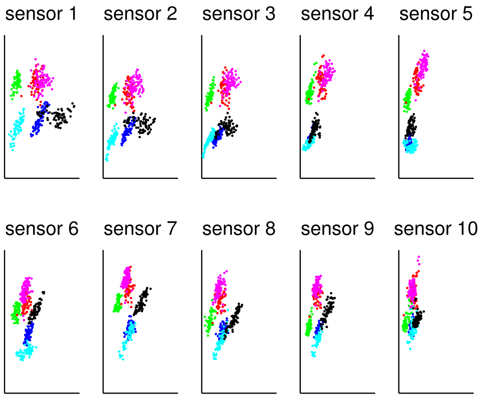

Fig. 2: Feature space of the fitness dataset.

B. Labeling

Annotations are required with supervised machine learning approaches. Acquiring labels is time consuming, costly and error prone.

We collected a dataset for the recognition of complex activities in sensor rich environments: the OPPORTUNITY dataset [4], [5]. Subjects performed a sequence of temporally unfolding and intertwined early morning situations: getting up, relaxing, preparing a coffee, drinking it, preparing a sandwich, eating it, cleaning up, and relaxing. We estimate that over 11000 and 17000 object and environment interactions occurred. We performed three types of annotations: *online annotation* by multiple experts (up to 5), *automatic annotation* of some activities by accurate ambient sensor (e.g. reed switches in the infrastructure), and *offline video-based.*

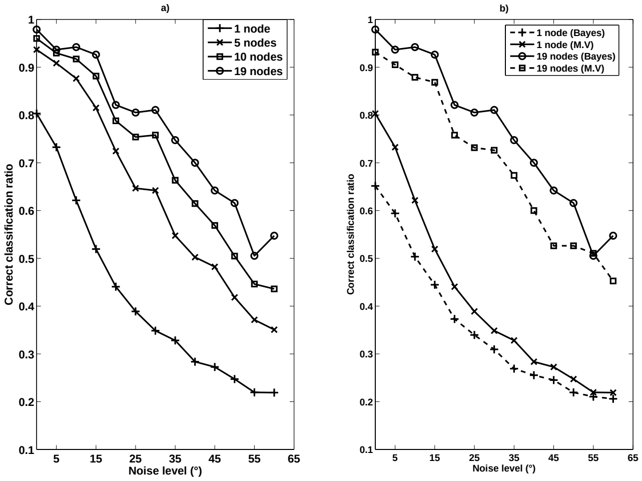

Fig. 3: a) Effect of rotational noise level on the fusion classification ratio for different cluster sizes.b) Comparison between naive Bayesian fusion and majority voting fusion for cluster size 1 and 19

Our experience showed that online annotation is error prone when subjects freely behave in the environment. Besides, during long and tedious recordings the experts may have difficulties concentrating. This results in missing labels, wrong labeling, or wrong label boundaries.

For some activities, however, online labeling is without alternative. We investigated the recognition of reading activity from eye movements in mobile daily life settings [6]. The experimental scenario involved participants reading text during different types of locomotion, including sitting, standing, and walking. Although reading involves characteristic eye movement patterns, online labeling was required to also catch the subtle and only short reading segments. These segments were particularly common during walking as participants regularly interrupted reading to check the way ahead. Labeling reading activity turned out to be challenging. A wearable video-based eye tracker would probably have allowed for the most accurate labels. Such systems require to wear headgear and additional equipment to process the video streams. To minimize the influence of labeling on the natural behavior of the participants, we labeled using the Nintendo Wii wireless controller. This approach turned out to be particularly useful as it allowed us to label robustly and fast, which resulted in the best quality of labels possible without distracting the participants too much. Although we still didn't achieve perfect ground truth annotation, with a view to a future realistic application scenario we decided to use the labels without manual tuning.

Automatic annotations are a bit more accurate, however it is also possible to have mislabeled data. For instance a person

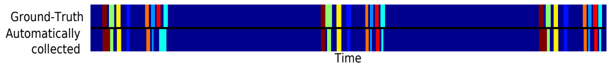

Fig. 4: Labels from offline video-based annotation (top) and automatic annotation from ambient infrastructure (down). There is label boundaries jitter and and some labels do not match.

opens a drawer, but not sufficiently so that the reed switch signals that the drawer has been opened half-way: an observer would correctly label this. The automatic system would detect the drawer is "not closed" but never reaches the "half-open" position and may discard the label as an artifact.

In the OPPORTUNITY setup 3 VGA cameras filmed the scene. Offline video-based annotation is usually considered as the "ground truth", yet errors may still occur:

- Small objects: it was difficult from the floor-mounted camera to distinguish objects in the hands. People annotating the video could in most case identify the objects, but it some case this wasn't possible and labels were omitted - or in the worst case wrongly assigned.
- Label boundaries: intertwined activities are difficult to label accurately as there isn't a clear pause between activity primitives (e.g. reach, grasp). There is no common agreement on these transitions and different persons labeling the data may interpret the video footage differently.
- Rapidity of human activities: activities can be quite fast. Finding the start and end of an activity (e.g. object grasped or released) is also limited by the speed of the camera. With a 15-fps camera, a jitter of the label boundary of 66ms (1 frame) isn't surprising. Usually more is expected as ones needs to look at the image sequence to identify activity primitives.

Thus, the notion of "ground truth" must be taken with a grain of salt:

- Wrong labels may be assigned, or segments of uninteresting data may be mistakenly labeled (user error)
- Some labels may be missing (e.g. due to the impossibility to distinguish an object)
- The label boundary may not be aligned to the activity boundary. For multiple repetitions of the activity, the alignment is likely to vary (label boundary jitter).

In figure 6 we illustrate the effect of inaccurate labels on the recognition performance, and the size of the training set. Inaccurate labels affect the recognition accuracy, measured according to the accurate ground truth (see table 5). The figure shows that a larger training set may be beneficial as unbiased labeling errors may cancel out (e.g. the curves with 79% 72% label accuracy). In figure 4 we show the difference between user-annotated data and automatically annotated data, including jitter in label boundaries and label disagreements.

II. Dealing with sensor and labeling variability

The problems mentioned above are not unique to our experience. Making activity recognition insensitive to placement and orientation variability has been investigated by many groups. Kunze et al. looked at methods to automatically infer on-body sensor placement and orientation [7], [8] and proposed methods to mitigate the effect of these variations [9]. Lester et al. [10] placed accelerometers at the shoulder, the wrist and the waist of subjects performing several activities of daily life to train an activity recognition system to be independent to sensor placement. In our own work, we investigated unsupervised classifier self-calibration as a way to deal with unpredictable sensor placement variability [2], and genetic-programming approaches to the generation of placement independent features [11].

Label Accuracy	**Classification Accuracy**
100%	79%
79%	71%
72%	66%
34%	28%

Fig. 5: Classification accuracy function of label accuracies.

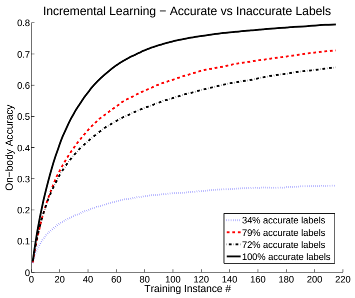

Fig. 6: Gesture recognition accuracy in function of the number of training instances and accuracy of the labels.

The problems of labeling variability or errors are also not unique to our experience. The work of Van Laerhoven et al. that focuses on extremely long recordings and relies on subjects labeling their activities a-posteriori also has to deal with inaccurate labels. This is mostly addressed by the use of efficient visualization and "smart" annotation tools, as well as the collection of sufficiently large amounts of data [12]. More generally, any use of experience-sampling is faced with similar problems. Classification with so-called "fuzzy labels" or "soft labels" is now actually an active area of research (see e.g. [13], [14]), especially with the use of crowd-sourced labeling of images and data on the web. In our own work [15] we investigate online learning, as a way to continuously adapt classifiers as new instances of activities arise, with opportunistic "automatic" labeling when ambient infrastructure is capable of providing activity labels. We showed this with *ContextCells*, an architecture for lifelong learning capable of sensing, online learning, classification and label exchange between cells.

III. Methodological and reporting recommendations

We outlined the problems caused by sensor placement and orientation variability, and the lack of accurate ground truth in the more complex activity recognition scenarios. Yet, in most of the literature published on activity recognition (including the work of our group) there is generally no mention of the consequences in terms of recognition performance.

Rather, one usually sees lots of efforts aiming at systematically placing sensors at the exact same locations in each repetition of an experiment. In real-world deployment of wearable activity recognition systems it is difficult, impractical or uncomfortable to achieve this. This approach is thus likely to create problems when the system is eventually deployed.

It is also not unseen to optimize label boundaries after recordings as a way to increase recognition accuracies. We don't believe this is legitimate as it only hides a problem that is also very likely to reappear when implementing activity recognition systems for real-world use. Instead, we suggest to put efforts in developing methods that are insensitive to - or are able to cope with - these kinds of variabilities. By using or developing methods robust to label variability or errors, one benefits from reduced labeling effort (and label 'tuning'), and overall reduced costs and faster 'time to market'.

Therefore, we suggest that the following points should be mentioned in publications w.r.t. sensor configurations:

- There should be a characterization of the influence of sensor variability - in particular orientation and placement - on the recognition performance.
- There should be a mention of the strategy used to mitigate the influence of placement/orientation variability (if any).
- If the system is insensitive to this variability, it is also interesting to outline the reasons for this.

With respect to labeling, we recommend:

- In general, ground truth annotations should not be tuned afterwards as this artificially increases the recognition performance. If they had to be tuned, this should be justified by the authors and a performance comparison based on the raw and tuned labels should be provided to show the influence of the tuning.
- There should be a characterization of the effect that the annotation quality has on recognition accuracy, as it indicates the robustness of the methods used.
- It should be indicated whether annotations during the experiment influenced the natural behavior of the participants. Otherwise, authors should explain how such an influence was prevented.
- There should be mention of the characteristics of the machine learning algorithms (or other parts of the recognition chain) that makes it robust to label quality variability.
- We suggest to also investigate new methods that are robust to label variability, as it may reduce cost and time associated with data labeling.
- Or there should be mention of the characteristics of the dataset that make it immune to the above problems (e.g. a very large dataset).

Obviously, addressing this in all publications is time consuming. It may not be meaningful or desired for articles presenting new methods or new application scenarios. However, for systems eventually designed for "real-world use" this is an aspect on which we invite the community to reflect on.

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

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

[3] P. Zappi, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and Tröster, "Activity recognition from on-body sensors by classifier fusion: Sensor scalability and robustness," in *3rd Int. Conf. on Intelligent Sensors, Sensor Networks, and Information Processing*, 2007.

[4] P. Lukowicz *et al.*, "Recording a complex, multi modal activity data set for context recognition," in *Workshop on Context-Systems Design, Evaluation and Optimisation, 23rd Int. Conf. on Architecture of Computing Systems*, 2010.

[5] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha, J. Doppler, C. Holzmann, M. Kurz, G. Holl, R. Chavarriaga, M. Creatura, and J. del R. Millán, "Collecting complex activity data sets in highly rich networked sensor environments," in *Submitted to the 7th Int. Conf. on Networked Sensing Systems*, 2010.

[6] A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster, "Robust Recognition of Reading Activity in Transit Using Wearable Electrooculography," in *Proceedings of the 6th International Conference on Pervasive Computing (Pervasive 2008)*. Springer, May 2008, pp. 19–37.

[7] K. Kunze, P. Lukowicz, H. Junker, and G. Troester, "Where am i: Recognizing on-body positions of wearable sensors," *LOCA'04: International Workshop on Locationand Context- . . .*, Jan 2005.

[8] K. Kunze, P. Lukowicz, K. Partridge, and B. Begole, "Which way am i facing: Inferring horizontal device orientation from an accelerometer signal," in *Proc. of Int. Symp. on Wearable Computers (ISWC)*. IEEE Press, 2009, pp. 149–150.

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

[10] J. Lester, T. Choudhury, and G. Borriello, "A practical approach to recognizing physical activities," in *Lecture Notes in Computer Science : Pervasive Computing*, 2006, pp. 1–16.

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

[12] K. Van Laerhoven and E. Berlin, "When did this happen? efficient subsequence representation and matching for wearable activity data," in *Proc. of the 12th Int. Symposium on Wearable Computers (ISWC)*, 2009, pp. 101–104.

[13] J. Lawry, J. W. Hall, and R. Bovey, "Fusion of expert and learnt knowledge in a framework of fuzzy labels," *International Journal of Approximate Reasoning*, vol. 36, pp. 151–198, 2004.

[14] C. Thiel, "Classification on soft labels is robust against label noise," in *Proc. Int. Conf. Knowledge-Based and Intelligent Information & Engineering Systems (KES), Part I*, 2008, pp. 65–73.

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