2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress

Human activity recognition of continuous data using Hidden Markov Models and the aspect of including discrete data.

Sarah Fallmann
Health & Environment Department
AIT Austrian Institute of Technology GmbH
Wiener Neustadt, Austria
sarah.fallmann.fl@ait.ac.at

Johannes Kropf
Health & Environment Department
AIT Austrian Institute of Technology GmbH
Wiener Neustadt, Austria
johannes.kropf@ait.ac.at

***Abstract*—The combination of discrete and continuous data during complex activity recognition is presented, as well as a concept to analyze continuous data. Accelerometer and gyrometer data gathered from a body worn sensor are analyzed by a continuous Hidden Markov Model (cHMM). This cHMM is evaluated through two comparative studies, producing better and comparable results. The recorded data differ from other datasets because of more complex activities, leading to a more realistic environment representation. The complicated part of this task are differences in single activities. On the one hand order diversity of the subactivities and on the other hand similar activities in one class, like food preparation include preparing breakfast and a three-course menu. The final output gives a hint to use body worn sensors in combination with binary sensors.**

***Keywords*-continuous Hidden Markov Model, MPU-9150 Sensor**

I. Introduction

Activity pattern recognition analysis based on continuous sensor data using the MPU-9150 sensor [1] is used to recognize daily activities in an everyday life environment. Activity recognition is currently a subject of intensive research, because of its importance in many fields.

The motivation of this work lies in specific in the growing generation of older adults, and the need to provide them a secure and appropriate living standard. The demographic changes lead to more people suffering from Alzheimer's and Parkinson's disease. The challenge of the increasing number of dementia patients can be approached by Ambient Assisted Living Technologies like activity recognition, to eliminate or make some tasks of care givers easier. This includes, among other things, sensors controlling kitchen appliances like stoves and guaranteeing the appropriate usage, due to activity recognition.

For Ambient Assisted Living (AAL), a system is demanded which can be adapted to different living conditions and is easily usable. The users of the system are primary older adults, but also care givers, who might use services from their own home. AAL technologies makes it easier to cope with impairments, monitor personal activity, ensure the security of older adults and reduce social isolation.[2]

The work focuses on the recognition of complex daily activities like tooth brushing, dinner preparation and changing clothes, different to many other datasets which only contain very simple actions[3] like drinking. A supervised classification algorithm, namely continuous Hidden Markov Model (cHMM), is used to detect different complex daily activities.

II. Related Work

There are many research studies over human activity recognition in different settings.[4], [5], [6], [7] Most of these works are based on acceleration data and tries to recognize daily activities like [4], [5], [7]. The main difference between the works lies in the choice of parameters in the different steps of recognition, meaning preprocessing, feature extraction, and finally training and classification.

Each study uses different sample frequencies during preprocessing. Bao *et al.*[4] used a sample frequency of 76.25Hz, Ravi *et al.*[5] and Shoaib *et al.*[6] used 50Hz. To get a hint which frequency is accurate in daily activity recognition, but still doesn't need too much memory, Khusainov *et al.*[7] compared different sampling rates and inferred that most of the body movements are contained in frequency below 20Hz.

Bao *et al.*[4] uses mean, energy, frequency-domain entropy, and correlation features and analyzed those via decision table, instance-based learning, decision tree (C4.5), and Naive-Bayes classifiers. They analyzed twenty activities like eating, bicycling or reading.

In [5] mean, standard deviation, energy and correlation are used with base-level classifiers like decision tables, k-nearest neighbours or Naive Bayes and meta-level classifiers like Boosting. With these classifiers activities like standing, walking, running, climbing up/down stairs, sit-ups, vacuuming and brushing teeth are recognized without noise filtering.

978-1-5090-2771-2/16 $31.00 © 2016 IEEE
DOI 10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.139

Shoaib *et al.*[6] record a combination of accelerometer, gyrometer and magnetometer data from a smartphone sensor and later six different activities with seven classifiers are analyzed. Shoaib *et al.*[6] show that the combination of accelerometer and gyrometer completes the system and gives better results during physical activity recognition. The feature calculation is kept as simple as possible with two time domain features. They handle four dimensions x, y, z and the magnitude $\sqrt{x^2 + y^2 + z^2}$ and compute mean and standard deviation.

In [8] six activities are recorded from a group of 30 persons with a sampling rate of 50Hz. All activities were performed twice with a smartphone on the waste recording the accelerometer and gyrometer data. They calculated a 561 features and experimented with a multiclass Support Vector Machine (MC-SVM), showing an overall accuracy of 96%.[8]

This work is mainly based on Bulling *et al.* [9], where body-worn accelerometer and gyrometer data are recorded to detect hand gestures. They recorded 12 activities and inbetween non-specific activities, so called 'NULL'-class. Data from two persons with three sensors placed on their arms in different heights are gathered. The sensor are placed on top of the right hand, outer side of the right lower and upper arm. The data comes from a three-axis accelerometer and a two-axis gyrometer, both recording annotated motion data at a sampling rate of 32Hz.[9]

In Chen *et al.* [10] new activities are tried to be learned from unlabeled data with data-mining methods to complete the activity recognition model. In this study an extra cHMM is used to detect subactivities and can also classify new activities in the 'NULL'-class.

Wearable sensors on their own are not suitable to monitor multiple interaction with the environment. In contrast, dense sensing based activity monitoring can, which records movements within a home, among others with binary sensors.[3] This study concentrates later on the combination of both data, discrete and continuous.

III. Activity recognition

The activity recognition consists the following steps: Sensor placement, data recording, preprocessing, feature extraction, training and classification with a cHMM.

A. Data

The InvenSense MotionFit^{TM} Software Development Kit is used to record data. The MPU-9150 is a nine-axis MotionTracking device optimized to fulfill the purposes for wearable sensor applications.[1]
The sensor is placed on the left hand wrist, with which the daily activities are mostly executed in a $59m^2$ flat. In table I the common daily activities, which are recorded in this study with a sampling frequency of 50Hz, are displayed. The activities are saved with their 3-axis accelerometer and 3-axis gyrometer data, which are widely used wearable sensors.[3] Inbetween all activities a 'NULL'-activity is performed, which consists of preparing the next activity and closing the preceding activity. The data gathering extends over days in many small sessions, which are finally put together to one dataset.

Table I
List of recorded daily activities

labels	**activities**	
1	NULL	
2	comb hair	
3	wash face	
4	wash hands	
5	brush teeth	electric/ non electric
6	make bed	
7	change clothes	
8	put blinds up/down	
9	prepare food	
10	eat with	folk/ spoon/ chopsticks
11	open/close window	
12	read newspaper/book	
13	putting shoes on	
14	drink from/with	straw/ mug/ cup

Data segmentation is needed to identify the segments of the data stream containing information about activities.[9] This is done by annotation during recording, saving information over start and duration of activities.

B. Preprocessing

Only data which are recorded more than once are used for analysis. Either by cutting out 'read newspaper/read book', 'putting shoes on' and 'drink from/with straw/mug/cup' from the whole dataset, with 'NULL'-classes between the activities remaining. Or by redefining the labels of those classes to the label of the 'NULL'-class. Other activities are put together to one: 'Tooth brushing electric' and 'Tooth brushing non electric' get label 5, 'Eat with folk', 'Eat with spoon' and 'Eat with chopsticks' are assigned to label 10. Each sensor records data in three dimensions and a fourth dimension describing the magnitude $\sqrt{x^2 + y^2 + z^2}$ is added.

Noise and artifacts are disturbances which can corrupt the human activity recognition, hence are reduced by common filters.[9] A Median filter and a 3rd order low-pass Butterworth filter are tested, also used among others in [8]. The 3rd order low-pass Butterworth filter has a cutoff frequency of 20Hz. This rate is sufficient, as the frequency of body motions is 99% below 15Hz.[8]

The application of the median filter causes a smoothing of the data, compare figure 1. Butterworth filters are used to cut high frequencies. The functionality of a third order low-pass Butterworth filter described above can be seen in figure 2. The original data is displayed in blue and the filtered data is displayed in red.

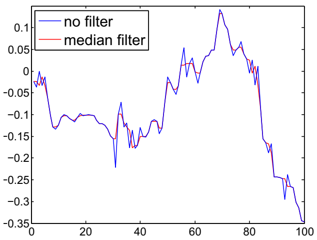

Figure 1. Median filter

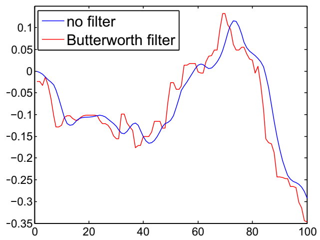

Figure 2. Butterworth filter

C. Feature Mapping

Features are calculated for each annotated activity with a shifted window sized 50, containing 50 data vectors, and an overlap of 50%, which is the most significant value for overlap in previous works.[11], [4] The mean, standard deviation, correlation [4], [5], energy [4], [5] and frequency domain entropy [4] are calculated.

The features can be divided into time domain features and frequency domain features. Time domain features are mean, standard deviation and correlation. Frequency domain features are energy and entropy.[6] A periodic function in time is described with a direct current (DC) component. The DC component over the window is the mean value. Standard deviation is important for the reason of different range of values for different activities. Periodicity in the data is saved in the energy feature. Correlation between axes is useful to differentiate activities with translation in one dimension. As example, walking and stair climbing can be distinguished over correlation data.[4], [5]
It is important to use a minimum number of features that allow good performance and at the same time minimize computational costs and memory.[9] Experiments show that entropy leads to no improvement of the results, therefore the best combination of features are used, namely mean, standard deviation and correlation.

D. Training

For activity recognition a supervised model which needs to get trained before operating, namely cHMM, is used.[9] Therefore the data has to be split into training and test data. As some activities are not so common, it is not possible to divide the data in usual 20% test and 80% training data. One activity is cut out from each activity class, with the 'NULL'-class behind for the test dataset. The remaining part is used as training data. An example for the structure of training and test data can be seen in figure 3 and 4.

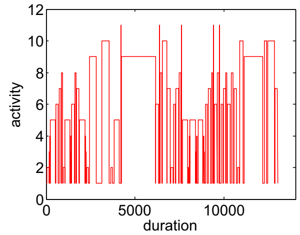

Figure 3. Training data

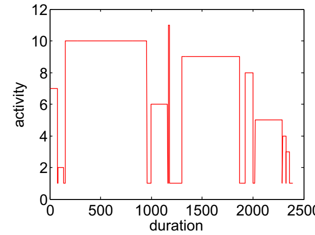

Figure 4. Test data

In cHMMs the model parameters $\theta = (\pi, A, B)$ are learned by minimizing the classification error.[9] In this work the transition matrix A and B, a list of pairs (μ, Σ) that define the distributions, can be calculated with the labeled training data. Only π has to be guessed.

E. Classification

The classification consists of two steps. The first one maps a set of class labels to each feature vector of the test data with corresponding scores. In the second step the scores are used to calculate the maximum score and take the corresponding class label y_i as the classification output.[9]

F. Performance Evaluation

The classification of the activities can be either correct 'True Positive' and 'True Negative' or wrong 'False Negative' and 'False Positive'. The performance metric which is used for this model is a confusion matrix, with accuracy, recall, specificity and precision. Beyond the confusion matrices and section IV-A, macro-averaged metrics are used.

The confusion matrix gives a breakdown of the misclassified activities. The rows show the instances in each actual activity class and the columns show the instances for each predicted activity class. The values in one row are the results from the comparison of all ground truth instances, from the actual class, to the predicted class labels.[9] In table II a confusion matrix can be seen, where the last column describes the recall values, the last row the precision values and the last box describes the accuracy.

IV. Validation

A. Bulling et al.[9]

In [9] data from 2 persons performing 12 activities are recorded with a 32Hz sampling rate: opening a window, closing a window, watering a plant, turning book pages, drinking from a bottle, cutting with a knife, chopping with a knife, stirring in a bowl, forehand, backhand and smash and a 'NULL'-class.[9] The sensor is placed on 3 positions on the right side: upper arm, lower arm, hand wrist.[9] The best outputs of Bulling [9] are constructed by using two features mean and standard deviation, for all 7 axes. This 7 axes come from the 3-axes accelerometer and 2-axes gyrometer, including one axis for each sensor, representing the magnitude.

The results from the cHMM are compared with the best hand wrist results from Bulling *et al.* [9] with the same overall precision and recall calculation. These circumstances are shown in figure 5, where the same dataset and features are used. The cHMM shows a higher recall value, caused by the different train and test data separation. In [9] the precision and recall lies by 87.2% and 55,1%, in contrast to 96.51% and 72.81% in this study. This states that the used cHMM is accurate.

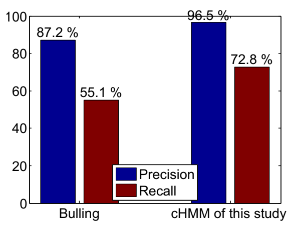

Figure 5. Precision and recall for sensor data from Bulling *et al.* [9] and this study.

B. Anguita et al.[8]

Anguita *et al.* gathered data from 30 volunteers via a Galaxy S II. smartphone on the waist, which followed a defined protocol of activities, consists of standing, sitting, laying down, walking, walking downstairs and upstairs. Accelerometer and gyrometer data are recorded with a sampling rate of 50Hz and 5 seconds break between two activities.[8]

Only a part of the 561 features vector of the provided test and training set is picked for evaluating the cHMM, meaning 107 features which are not highly correlated to each other. The feature data is already noise reduced by a median filter and a 3rd order low-pass Butterworth filter with a 20Hz cutoff frequency.[8] In table II the confusion matrix of the cHMM is depicted. In table III the results of Anguita *et al.*[8] using MC-SVM are reproduced. The smaller accuracy can be attributed to the more complex MC-SVM model in [8]. The precision of the class 'walking' and the recall of 'sitting' is even better compared to MC-SVM.

V. Experiments

Before section V-C the activities which occur only once in the recording period are relabeled as 'NULL'-class. After section V-C the once recorded classes are cut out of the whole dataset, leading to one percent improvement. In real environmental applications this makes a small difference and therefore is not necessary.

Table II
Confusion matrix of cHMM

Walking	W.Upstairs	W.Downstairs	Sitting	Standing	Laying Down	
442	0	54	0	0	0	89.11
0	439	32	0	0	0	93.21
0	10	410	0	0	0	97.62
1	0	1	472	7	10	96.13
1	2	8	31	480	10	90.23
2	0	0	70	0	465	86.59
99.10	97.34	81.19	82.37	98.56	95.88	91.89

Table III
Confusion matrix of MC-SVM [8]

Walking	W.Upstairs	W.Downstairs	Sitting	Standing	Laying Down	
492	1	3	0	0	0	99.12
18	451	2	0	0	0	95.75
4	6	410	0	0	0	97.62
0	2	0	432	57	0	87.98
0	0	0	14	518	0	97.37
0	0	0	0	0	537	100
95.72	98.04	98.80	96.86	90.09	100	96

A. Training and test sets

Different sort of training and test data partitions are analyzed. For example, the test data includes the third repetition of each single activity class and takes either the 'NULL'-class behind or in front of each cut activity. Another approach uses the second repetitions. The Viterbi path (blue) with the original labeled path (red) is depicted in figure 6, 7. Out of the table IV, illustrating the changes in accuracy, specificity and sensitivity, the conclusion can be drawn, that the 3rd activities with 'NULL'-class behind implies the best result.

Table IV
Accuracy, specificity and sensitivity for different sets

Experiment	accuracy	specificity	sensitivity
3rd back	80.24	97.92	72.18
3rd front	79.84	97.82	71.91
2nd back	63.99	96.23	66.94
2nd front	64.67	96.38	68.06

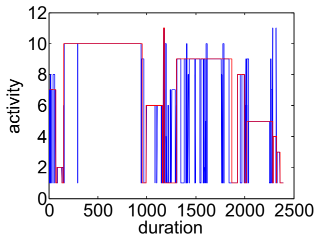

Figure 6. 3rd activities with 'NULL'-class behind

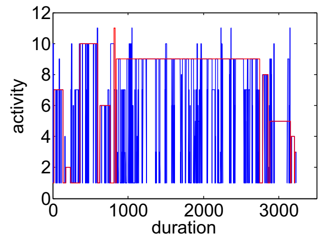

Figure 7. 2nd activities with 'NULL'-class in front

Different feature combinations are compared, leading to the most appropriate combination of mean, variance, correlation without the magnitude for the 3rd back dataset.

B. Filters

A median filter and a 3rd order low pass Butterworth filter with a corner frequency of 20Hz is tested to remove noise, based on Anguita *et al.* [8].

For the best output of section V-A, accuracy, specificity and sensitivity of filtered and non-filtered data are represented in table V. The results get worse, therefore filters seem unnecessary for this dataset.

Table V
Accuracy, specificity and sensitivity for filtered and non filtered data

Experiment	accuracy	specificity	sensitivity
no filter	80.24	97.92	72.18
filter	79.53	97.86	70.40

C. Accelerometer/Gyrometer

Accelerometer and gyrometer data on their own are analyzed and compared with results of Bulling *et al.* [9].

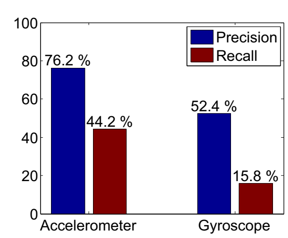

Figure 8. Precision and Recall by Bulling *et al.* [9]

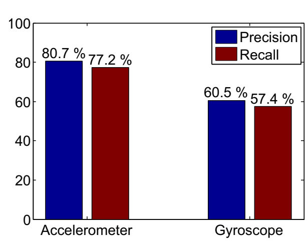

Figure 9. Precision and Recall of this study

The single accelerometer dataset is more accurate than the single gyrometer dataset. These results coincide with those of Bulling *et al.* [9] using different data. These circumstances are illustrated in figures 8 and 9, where precision and recall are symbolized as blue and red bars. The outputs of this study reach higher level of precision and recall, even though the dataset is smaller and has more complex activities. The combination of accelerometer and gyrometer data have slightly worse precision and recall values, therefore the conclusion can be drawn that gyrometer data is unnecessary in this case.

D. Extra cHMM

For each often misclassified activity-class, a cHMM is applied to divide the activities in subactivities. The number of subactivities depends on the number of states in the cHMM. Hence, an iteration is done, constructing a 2-to-5-state cHMM, choosing the cHMM with the highest accuracy. For example the improvement by using the divided 'Tooth brushing' class comes from the case that electric tooth-brushing has a higher frequency.

The extra cHMM model is very sensitive. During the experiments only the parameters, 'k'-fold cross-validation and number of activity classes divided, are considered. Change in tolerance and of maximal iterations within the cHMM effects the outcome as well. The tolerance is always set to $1e^{-5}$ and the maximal iteration is set to 10.

In figures 10 the basic cHMM is shown, in contrast to figure 11, where the results for including subactivities for tooth brushing, putting blinds up/down and prepare food are displayed, symbolized as Viterbi path (blue) correlating with the original labeled path (red). The results with subactivities show a better fit to the original path with an accuracy of 83.84 in contrast to the basic cHMM accuracy of 81.21.

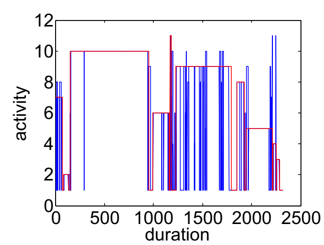

Figure 10. Basic cHMM

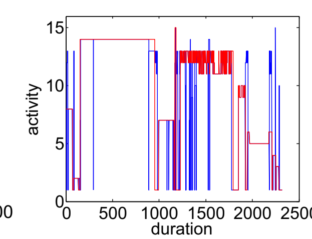

Figure 11. Extra cHMM

E. Continuous/Discrete data

Test and training data get an additional column, consisting of the room number, describing the room where the activities are performed. 'NULL'-class activities become half the rooms label from the previous activity and half the room label from the following activity. The process represents the usage of smart home binary sensors in combination with body-worn sensors. As can be seen in table VI, the accuracy improves about 8% and sensitivity about 15%, while specificity stays nearly the same. This is a great improvement justifying the effort of collecting both data, continuous and discrete.

Table VI
Accuracy, specificity and sensitivity for continuous and continuous & discrete data

Experiment	accuracy	specificity	sensitivity
continuous	81.21	98.02	72.63
continuous & discrete	88.75	98.85	88.09

The better output can also be seen in the comparison of the Viterbi path (blue) and the original labeled path (red) in figure 12 and 13. Figure 12 shows results with continuous data and figure 13 with the combined dataset.

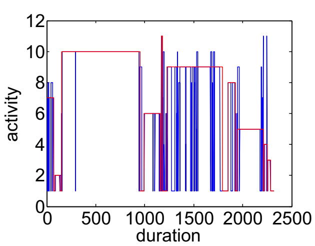

Figure 12. Continuous data

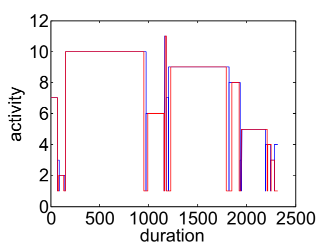

Figure 13. Con. & discrete data

VI. Conclusion and Outlook

Human activity recognition in AAL using a 3-axis gyrometer and a 3-axis accelerometer is performed. This raw data are preprocessed and split into test and training datasets. Later on features are extracted used to construct a cHMM.

A validation of the cHMM model is done with provided data from Bulling *et al.* [9] and Anguita *et al.* [8]. Afterwards different experiments were accomplished. Using only accelerometer data with mean, variance and correlation leads to the best results. The conclusion is that gyrometer data are not necessary and filters do not really contribute to significant improvement. Combining discrete and continuous data considerably improves the results, which is the most important outcome.

Research should focus on the combination of discrete data from binary sensors and continuous data from wearable sensors. This will lead to more robust and trustable models. A broader consideration, meaning a bigger dataset with more activities and people included, would lead to results which allow to imply more general statements.

References

[1] InvenSense Inc., "MotionFit™ Wireless Developer Kit," 2012, accessed: 2015-02-06. [Online]. Available: http://invensense.com/mems/gyro/documents/PB-MPU-9150IMF$%$20MotionFit$%$20Wireless$%$20Developer$%20$Kit$%$20Product$%$20Brief.pdf

[2] G. Van Den Broek, F. Cavallo, and C. Wehrmann, *AALIANCE Ambient Assisted Living Roadmap*. Amsterdam, The Netherlands, The Netherlands: IOS Press, 2010.

[3] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 790–808, Nov 2012.

[4] L. Bao and S. S. Intille, "Activity Recognition from User-Annotated Acceleration Data." in *Pervasive*, ser. Lecture Notes in Computer Science, A. Ferscha and F. Mattern, Eds., vol. 3001. Springer, 2004, pp. 1–17.

[5] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity Recognition from Accelerometer Data," in *Proceedings of the 17th Conference on Innovative Applications of Artificial Intelligence - Volume 3*, ser. IAAI'05. AAAI Press, 2005, pp. 1541–1546.

[6] M. Shoaib, H. Scholten, and P. Havinga, "Towards physical activity recognition using smartphone sensors," in *10th IEEE International Conference on Ubiquitous Intelligence and Computing, UIC 2013*. Los Alamitos, CA, USA: IEEE Computer Society, December 2013, pp. 80–87.

[7] R. Khusainov, D. Azzi, I. E. Achumba, and S. D. Bersch, "Real-Time Human Ambulation, Activity, and Physiological Monitoring: Taxonomy of Issues, Techniques, Applications, Challenges and Limitations," *Sensors*, vol. 13, no. 10, pp. 12 852–12 902, 2013.

[8] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A Public Domain Dataset for Human Activity Recognition using Smartphones," in *21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013*, 2013, bruges, Belgium.

[9] A. Bulling, U. Blanke, and B. Schiele, "A Tutorial on Human Activity Recognition Using Body-worn Inertial Sensors," *ACM Computing Surveys*, vol. 46, no. 3, pp. 33:1–33:33, 2014.

[10] L. Chen, C. Nugent, and J. Rafferty, "Ontology-based activity recognition framework and services," in *Proceedings of International Conference on Information Integration and Web-based Applications & Services*, ser. IIWAS '13. New York, NY, USA: ACM, 2013, pp. 463:463–463:469.

[11] K. Van Laerhoven and O. Cakmakci, "What Shall We Teach Our Pants?" in *ISWC '00: Proceedings of the 4th IEEE International Symposium on Wearable Computers*, IEEE Computer Society. Washington, DC, USA: IEEE Computer Society, 2000, p. 77.