Department of
**Information Engineering
and Computer Science** **DISI**

UNIVERSITY
OF TRENTO - Italy

DISI - Via Sommarive 5 - 38123 Povo - Trento (Italy)
http://disi.unitn.it

RECOGNIZING HOSPITAL CARE ACTIVITIES WITH A COAT POCKET WORN SMARTPHONE

Gernot Bahle, Agnes Gruenerbl, Paul Lukowicz, Mattia Zeni, Enrico Bignotti, Fausto Giunchiglia

Date (2014-10)

Technical Report # DISI-14-015

RECOGNIZING HOSPITAL CARE ACTIVITIES WITH A COAT POCKET WORN SMARTPHONE

Gernot Bahle, Agnes Gruenerbl, Paul Lukowicz, Mattia Zeni, Enrico Bignotti, Fausto Giunchiglia

Embedded Intelligence, German Research Center for Artificial Intelligence
Kaiserslautern, Germany
{author1, author2 and author3}@dfki.de

Department of Information Engineering and Computer Science,
University of Trento, via Sommarive 14 I-38123, Povo, Trento, Italy
{author4, author5 and author6}@disi.unitn.it

Abstract. In this work, we show how a smart-phone worn unobtrusively in a nurses coat pocket can be used to document the patient care activities performed during a regular morning routine. The main contribution is to show how, taking into account certain domain specific boundary conditions, a single sensor node worn in such an (from the sensing point of view) unfavorable location can still recognize complex, sometimes subtle activities. We evaluate our approach in a large real life dataset from day to day hospital operation. In total, 4 runs of patient care per day were collected for 14 days at a geriatric ward and annotated in high detail by following the performing nurses for the entire duration. This amounts to over 800 hours of sensor data including acceleration, gyroscope, compass, wifi and sound annotated with groundtruth at less than 1min resolution.

1 Introduction

The main goal of this work was to see how well activities relevant for nursing documentation in a hospital can be recog- nized within boundary conditions found in

an average hospital. While a large body of work on activity recognition already exists, it is often limited by either an artificial, lab setting, or otherwise only deals with somewhat simple activities like modes of locomotion. Furthermore, much work also requires extensive instrumentation of either the environment or the sub- jects or even both. Given these considerations, it is understand- able that especially in a hospital setting, activity recognition and the applications it can support or enable is still very much an active and difficult research topic. One of the tasks that health care professionals usually face when treating patients is an exact and meticulous need for documentation, for various reasons like treatment monitoring, quality of care and legal safeguarding. Unfortunately, this is also a very time-intensive process. Given the high workload nurses face, reducing time spent here would be beneficial. This article presents a study aimed at laying the groundwork for exactly that. Especially, this means trying to recognize complex activities in a real-life setting under real-life working conditions.

A. Related Work

The usage of wearable and pervasive technology in health- care is not a new field any longer. It has already been explored in numerous publications. Overviews include [1] or [2]. The usefulness of context in wearable computing in general is e.g. discussed in [3]. Kunze et al. [4] describe a quanti- tative Wizard-of-Oz study of using context recognition for a wearable maintenance assistance system. Current research about activity recognition reveals a wide range of approaches. Much work was done on the use of inertial sensors such as accelerometers, gyroscopes or magnetometers. [5], [6].

Activity Recognition in real-life often focuses on recogniz- ing simple activities like "walking", "sitting", "standing" or "running" [7] or gym-kind exercises (see [8] or [9]). For more complex activities like workshop or maintenance work, often a large number of sensors is deployed to reach sufficiently accurate results (see [10], [11]). If the setting becomes more complex, additional information sources beyond sensor infor- mation are required. Here, approaches like hidden Markov models (HMMs) [12] and conditional random fields (CRFs) [13] are utilized to represent model- based information (e.g. about the sequence of events). In [14], relational Markov networks are used to represent a hierarchical activity model to determine location based activities from GPS traces. Zinnen et. al [15] evaluate sensor-based vs. model- based activity recogni- tion. Other approaches try to transfer learned knowledge from one application to others thereby reducing the required training data for new applications [16].

In health care pervasive computing and context recogni- tion mainly includes context aware systems, which provide information for care documentation. A future application is introduced by Agarwal et al. [17] who describe a prototype context aware perioperative information system to capture and interpret data in an operating

room of the future. The captured data is used to construct the context of the surgical procedure and detect medically significant events. Such events, and other state information, are used to automatically construct an electronic medical encounter record. In [18] a study is introduced, which implemented a nursing support system in a series of laboratory experiments under simulated conditions, where clinical nurses were asked to perform assigned nursing tasks and simulated patients were used to make the environ- ment realistic. Another system introduced by [19] describes a wireless sensor network system for supporting context- awareness of nursing activities in hospitals.

B. Contributions

This work describes a study set in a real hospital, done during regular working hours with live patient care. Our main contributions are:

- Recording of a large dataset under real life conditions during the usual operation of a regular hospital ward containing over 800 hours of sensor data; this amounts to over 130 care routines for multiple patients, performed by multiple nurses, exactly labelled by trailing researchers

- Recognition of a complex set of activities relevant for nursing documentation using only a standard (and thus cheap, readily available) smartphone worn unobtrusively in a coat pocket (and thus suited for daily work). It is important to point out that no obtrusive, extensive instru- mentation or special hardware was used or necessary.

- Further augmentation of recognition accuracy by includ- ing best practice information supplied by a "nursing workflow" derived from documentation and observation. We believe this approach may also be of value to other researchers for different scenarios.

2 The Scenario

As we described in the Introduction our work was motivated by the request for an application assisting nursing documenta- tion and the need to develop a solution that can be deployed in an average hospital. From this three main constraints resulted:

1) The only sensor placement on which the nurses could agree was to just put it into the coat pocket (see figure 1). Not all had pockets in the trousers and strapping anything to the body was considered to be too much of a disruption

and potential source of injury when patients would hold on to a nurse. Furthermore, all other locations could easily expose the sensor to damage.
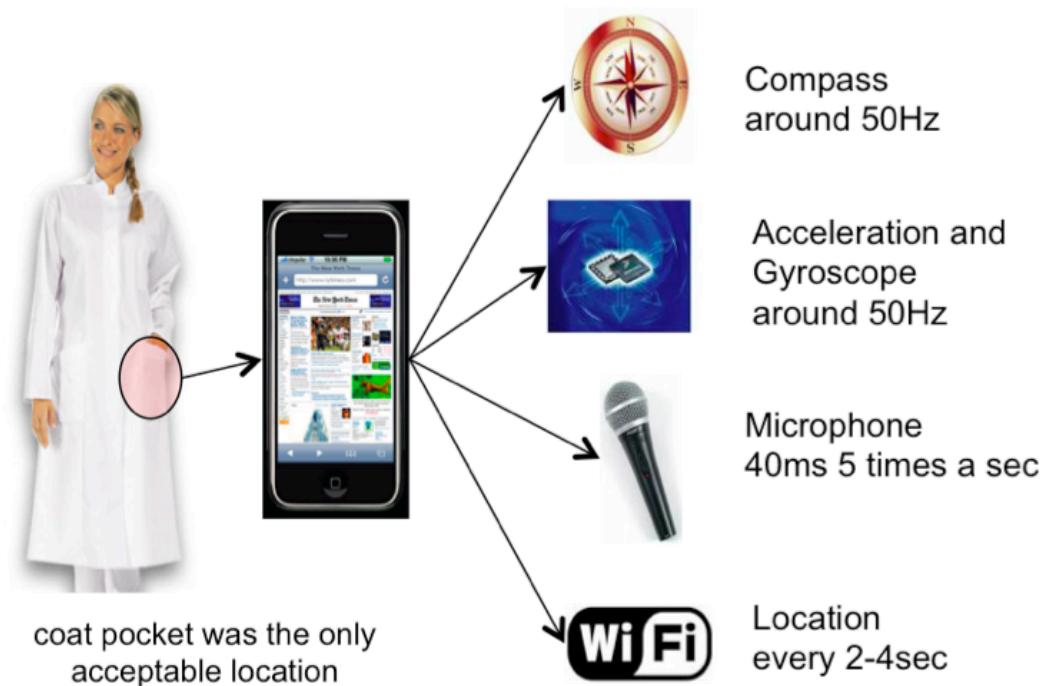

Fig 1. Sensor setup: simply a smart-phone (with its internal sensors) placed in the nurse's coat pocket

2) No videos were allowed and sound could only be recorded if it was cut into pieces and randomly mixed so that no speech could be retrieved.

3) We could follow the nurses and make all the notes we wanted, but questions and repetitions of activities for any sort of training were not allowed during the nursing process.

A. Real-life Data Capture of Morning Activities

In the hospital ward chosen, four nurses performed a morning hygiene routine for a few (2 to 3) patients each, including activities like washing, showering, drying,

dressing, making the bed, measuring pulse and blood pressure and so on. The experiment at the ward lasted for a total duration of 14 days, each with 4 recorded runs of the morning shift procedure with a total of 18 different nurses. Overall, more than 800 hours of sensor data were recorded. Each of these runs was slightly different, both because the nurses were free to chose the order of execution of some activities as well as the need to fit them to a patient's requirements. The platform used for sensing was a standard smart-phone, placed in the coat pocket of each nurse. Major constraints were invisibility towards patients and unobtrusiveness, as it was not permissible to obstruct or hinder the nurses in any way. The recorded sensor data was annotated by a researcher trailing each nurse, using Ipads running a labeling application developed for this study. This allowed annotations at less then 1min scale. To maintain the quiet atmosphere on the ward, only two of the four nurses work could be annotated every day. As a result, 30 sets of activity annotations (including app. 130 patient- and app. 120 nurse flow executions) were collected.

B. Building a Model

To provide a meaningful description of the sequence and nature of activities encountered, a hierarchical workflow model for the morning hygiene routine was created by analysis of best practice guidelines and the labels collected (see figure 2). Interestingly, although medical procedures are subject to strong restrictions in their execution, the ward does not provide detailed workflow models for the patient care procedures on their own. While well trained medical staff does not need this information, for our research these models and definitions were required. For our purpose, the model assembled con- tained activities performed and their possible sequences (e.g. dressing will only be performed after showering, which may, however, be optional). It was further enhanced with relevant context annotations and restrictions such as spatial limitations (i.e. a task can only be performed in one of several given locations) and context requirements like limitations on mode of locomotion of the task (e.g. not possible to measure pulse while walking).
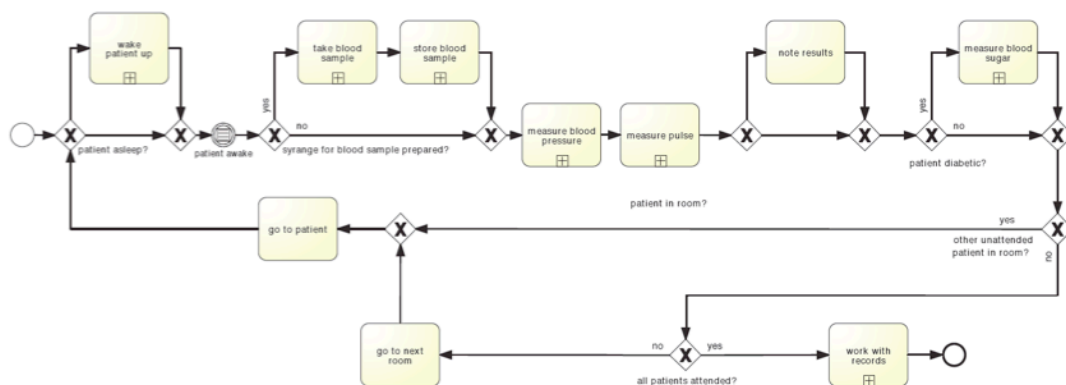
Fig 2. Nursing flow of the morning examination tasks

3 Approach

A. Constraints

Although the chosen smart-phone platform features 8 poten- tial sources for context relevant data in general scenarios, only 5 of these could be used in the hospital trial conducted: while Accelerometer, Gyroscope, Compass, Microphone and WiFi were used to record data, Bluetooth, Camera and GPS were not available as usage was generally not possible or would have required changes in the ward environment (i.e. installation of devices or markers in patient inhabited areas) which were not permitted.

The first difficulty in dealing with the sensor data was given by the nature of the experiment. Since data was collected during regular work at the ward, it was not possible to repeat actions or pause execution. Also, it was not possible to record video or unscrambled audio for labelling purposes. As a result, while the labelling information gained by following the nurses and noting their actions in our Ipad app is reasonably accurate, it is not accurate enough for standard supervised training methods. As an example, suppose showering was to be recognized. Further suppose this activity lasts for 5 minutes and the signal is divided into 5 second windows. Both acceleration and sound data will contain characteristic frames that would well represent showering. However, not only will there be others that only amount to background noise, even more important, the windows actually containing the relevant information may be (and in all likelihood are) different depending on sensing modality. As a consequence, direct fusion of all modalities at the feature level is not feasible, as this would mask relevant information by the noise of other sensors.

B. Recognition Approach

Keeping the above constraints in mind, we decided upon the following approach (also see figure 3)

Fig 3. Recognition architecture: leverageing flow information in order to enhance the recognition results

1. *Step 1*: Recognition on single sensors:: For both inertial sensors and sound, the signal was divided into 4 second windows. For inertial sensors, standard features were calculated on the (orientation invariant) norm, e.g. min, max, rms, frequency centroid, etc.. For sound, [20] served as tem- plate, with

features heavily frequency based (e.g. frequency centroid and bandwidth, band energy ratios, etc.). Frames containing no or little information were marked as 'null' and not considered further. Criteria included no movement for inertial sensors and silence for sound. On the remaining windows, a Bayes classifier was trained using data from 8 of the 20 annotated datasets. For every window, this resulted in a predicted activity or the label 'null' (nothing happening) if the most likely probability was still below a threshold. WiFi data was used in a fingerprinting approach. For all access points visible throughout the ward, signal strength was gathered and represented as an n-dimensional vector. Points close to one another then tended to exhibit closeness in this n-dimensional space, even if single components may fluctuate. Since wifi signal geometry is influenced by furniture, humans and environmental conditions, all of which can change, this approach only gives a rough idea of location (somewhere between individual rooms and and a group of rooms close to one another) and cannot be used to pinpoint a nurse's exact position.

2. *Step 2*: Combining sensor modalities:: As a next step, these single sensor, frame by frame results were combined and elevated to the event level. To that end, for each activity its average duration was calculated; afterwards, again for each activity, a sliding window of that duration was moved along the individual frame by frame results. All frames within were counted according to their classifier's confusion matrix. An example: assume that there are classes A and B, with our classifier recognizing A 90% as A and 10% as B. If an individual frame was then classified as A by said classifier, it would count 0.9 towards the 'A score' and 0.1 towards the 'B score'. The coarse positioning obtained by WiFi was used as a constraint; if an activity was not possible in the rough area provided by the location classifier, it was counted towards 'null'. Once all frames within the window were counted, a majority decision was performed, the winner (and its associated winning ratio) being noted down for that time interval. If subsequent windows contained the same activity, they were merged. If two activities collided (remember, there was a sliding window for each one), the one with the higher winning ratio was selected.

3. *Step 3*: Workflow Integration Level:: After the fusion step, domain specific knowledge codified into a best practice workflow was applied to further enhance recognition results. We would like to point out that this workflow only served as a template. The actual information was derived from the labels recorded during the study. This information was transformed into a finite state machine, with the transition probabilities calculated like this: suppose activity A was followed by B 10 times and by C 20 times. The edge (A, B) would then carry probability 33%. The resulting finite state machine

was then combined with the sequence of activities (each associated with a ratio) from the previous step. Starting from the first event, for each following event a new likelihood was calculated as a weighted product between its ratio from the fusion step and the transition probability of the finite state machine. The weighting itself was done by a sigmoid function; i.e. the higher a probability, the more influence on the result it had.

4. Results

A. Solely Sensor-based Results

The results of steps 1 and 2 described above (recognition through sensor fusion without the information about the work- flow) are shown in 4. The average precision and recall are 48.6% and 49.9%. Given the difficulty of the task, resulting from the combination of unsuitable sensor placement and complex activities, this is not unexpected. The results for in- dividual activities are also in line with expectations. Activities associated with a loud, characteristic sound are recognized well. This includes showering, hair drying and shaving. More subtle yet characteristic sounds are associated with blood pressure measurement (putting on the calf, pumping) or putting on gloves (when opening the plastic bag with the one way gloves). Activities such as dressing the patient and washing the patient in bed have a strong motion component with some sound and are also recognized fairly well. On the other hand activities where the motion is mostly restricted to hands with little sound such as changing a bandage or drying the patient are recognized very poorly. It is interesting to note that measuring pulse (manually with a finger on the wrist), which has neither sound nor motion is recognized better then may be expected. This is precisely due to the fact that there are little situations beyond pulse measurement where there is so little sound or motion.

B. Results with workflow information

The improvement that can be achieved when using the workflow information can be seen in Figure 5. The average precision and recall rise to 72.0% (plus 23.6 percentage points) and 72.4% (plus 22.5 percentage points). The improvement can be seen across all activities improving the "good ones" towards around 90% but even more dramatically raising the rates for the "bad ones" such as changing a bandage and drying the patient. How much the activities improve clearly depends on the amount of constraints imposed by the workflow. It is also interesting to notice, that the errors are mostly due to substitutions rather than insertions or deletions. Basically, this means that the system may be wrong about what activity is happening, but in almost all cases it will both notice correctly that some event is

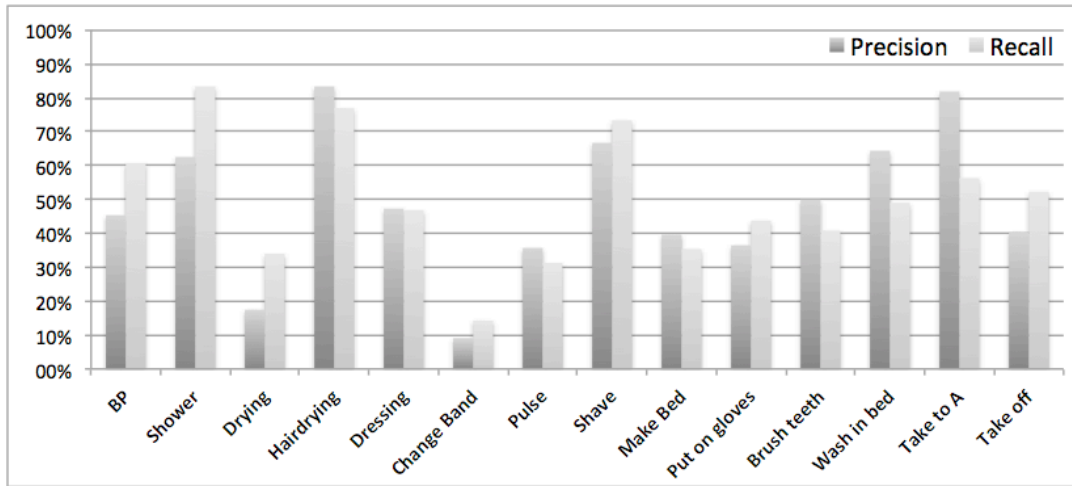occurring and at the same miss almost none of them.

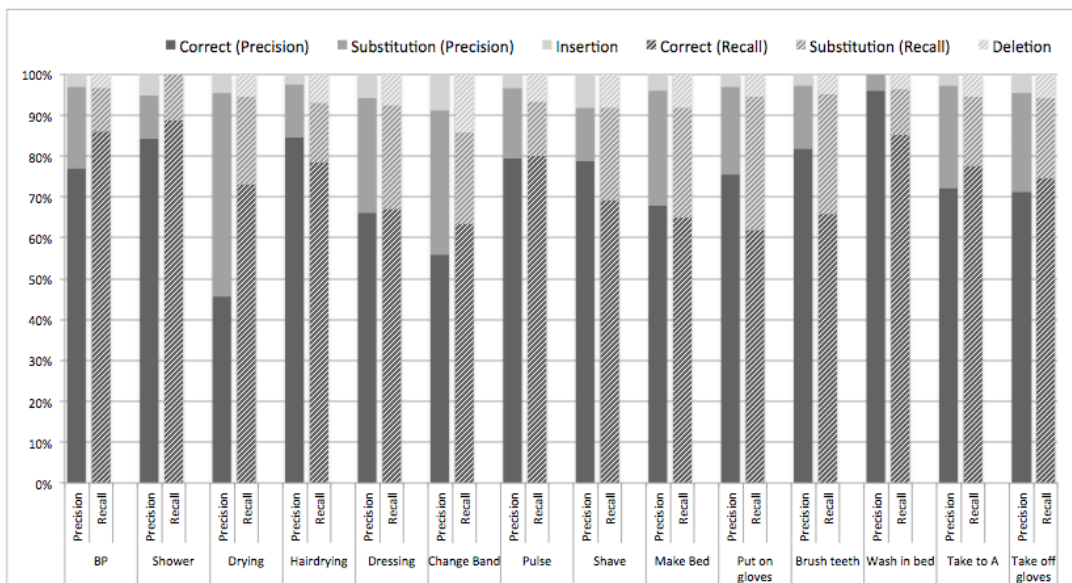**Fig 4**. Precision and Recall without semantic knowledge

Fig 5. Precision, Recall and Errors, using semantic knowledge

Given this, there is another interesting evaluation especially pertinent to context sensitive nursing documentation. The output of the classifier is a list of all possible

activities, ranked by their inferred probability (i.e. the system may think the event is 70% showering, 20%washing, etc.). If the output of the system is used for listing the most likely candidates for a nurse to pick as done in [21], it is already enough for the correct activity to be among the best two or three candidates. Given that metric, as shown in Figure 6, the correct activity is within the 2 best picks in 78.7% of cases and within the 3 best picks in 85.3%. The exact type of confusions are shown in Table I.
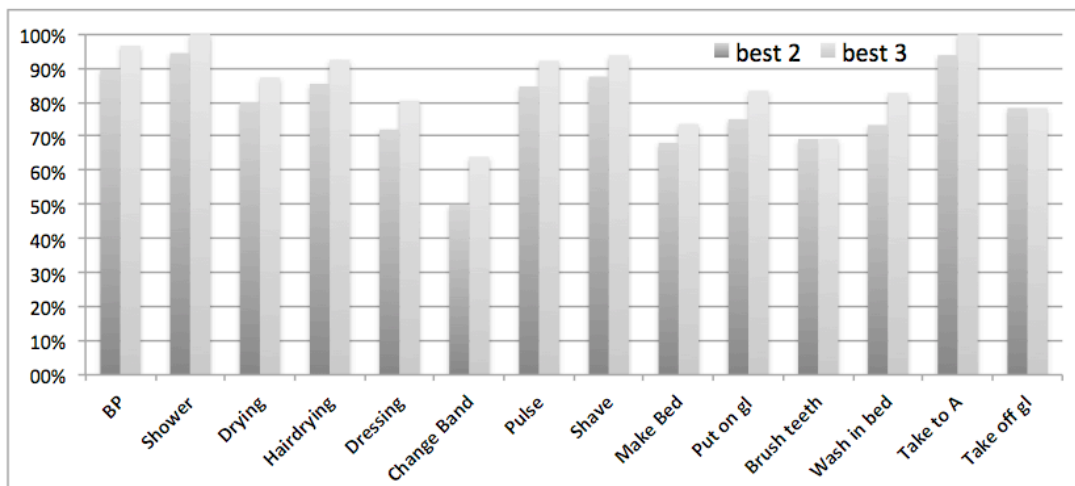

Fig 6. Percentage of the correct activity being among the best 2 or 3

	BP	Shower	Dry.	Hairdr.	Dress.	Ch. Band	Pulse	Shave	M. Bed	Put on gl	Br. teeth	W. i. bed	Take to A	Take off gl	NULL	Sum
BP	50								2						2	56
Shower		16	1									1			0	18
Drying			41		6	2						2		2	3	56
Hairdrying				33	3	1		1			1				3	42
Dressing			15	4	149	12						25			17	222
Change Band	1				2	11	1			2		2		2	5	26
Pulse	5						40			3					5	53
Shave			1					26	1			1		1	2	32
Make Bed	2				5		2	1	46	3		4	1	1	7	72
Put on gloves			3		2	2	3	1	2	57	2	3		5	7	87
Brush teeth			2				2		5		31	6		3	3	52
Wash in bed	3	2	24		43		2	4	9	4		248		4	16	359
Take to A									1	1			29	1	1	33
Take off gl.	2				2			1	3			1		35	3	47
NULL	2	1	4	1	13	4	3	1	5	4	1	8	0	2		47

Tab 1. CONFUSION MATRIX FOR WORKFLOW ENHANCED RECOGNITION

5. Conclusions and Discussion

Overall, this work shows that even under unfavorable sensor constraints, given by the deployment in a real world environ- ment such as a hospital, reasonable

precision and recall can be achieved when exploiting workflow concerns. While clearly, this approach is not applicable to all situations, nevertheless, especially in professional settings where the aim of activity recognition is documentation or context sensitive support, there are often strong workflow constraints. These are at the same time often the situations where practicability issues make the deployment of rich, well placed sensor configurations and the collection of adequate training data difficult.

Note that as described in [21] a documentation support system does not need perfect recognition. If we can reduce the documentation to a selection from a few alternatives and keep the number of deletions small then the process can already be greatly optimized. Thus the results presented here are of significant practical interest.

As a next step, we will investigate integrating low level data analysis in a closer and more generalizable way with a semantic high level representation. By high level representa- tion we mean a semantic model, that can be broadly defined as a representation of the scenario and the relations among its objects.

6. Acknowledgements

References

[1] P. Lukowicz, "Wearable computing and artificial intelligence for health- care applications." *Artificial Intelligence in Medicine*, vol. 42, no. 2, pp. 95–98, 2008.

[2] C. Orwat, A. Graefe, and T. Faulwasser, "Towards pervasive computing in health care– A literature review," *BMC Medical Informatics and Decision Making*, vol. 8, no. 1, p. 26, 2008.

[3] A. Schmidt, M. Beigl, and H. Gellersen, "There is more to context than location," *Computers and Graphics*, vol. 23, pp. 893–901, 1998.

[4] K. Kunze, F. Wagner, E. Kartal, E. Morales Kluge, and P. Lukowicz, "Does Context Matter?-A Quantitative Evaluation in a Real World Maintenance Scenario," *Pervasive Computing*, pp. 372–389, 2009.

[5] L.BaoandS.S.Intille,"Activityrecognitionfromuser-annotatedaccel- eration data," in

Pervasive Computing, ser. Lecture Notes in Computer Science, A. Ferscha and F. Mattern, Eds. Springer Berlin Heidelberg, 2004, pp. 1–17.

[6] J.Lester,T.Choudhury,N.Kern,G.Borriello,andB.Hannaford,"Ahy- brid discriminative/generative approach for modeling human activities," in In Proc. of the International Joint Conference on Artificial Intelligence (IJCAI, 2005, pp. 766–772.

[7] G.Bieber,J.Voskamp,andB.Urban,"Activityrecognitionforeveryday life on mobile phones," in Proceedings of the 5th International on Con- ferenceUniversal Access in Human-Computer Interaction. Springer- Verlag, 2009, pp. 289–296.

[8] M. Muehlbauer, G. Bahle, and P. Lukowicz, "What can an arm holster worn smart phone do for activity recognition?" in ISWC, 2011, pp. 79– 82.

[9] C. Seeger, A. Buchmann, and K. Van Laerhoven, "myhealthassistant: A phone-based body sensor network that captures the wearer's exercises throughout the day," in The 6th International Conference on Body Area Networks. Beijing, China: ACM Press, 11 2011.

[10] G. Ogris, T. Stiefmeier, P. Lukowicz, and G. Trster, "Using a complex multi-modal on-body sensor system for activity spotting." in ISWC. IEEE, 2008, pp. 55–62.

[11] A. Zinnen, C. Wojek, and B. Schiele, "Multi activity recognition based on bodymodel-derived primitives." in LoCA, ser. Lecture Notes in Computer Science, vol. 5561. Springer, 2009, pp. 1–18.

[12] T. Starner, J. Weaver, and A. Pentland, "Real-time american sign language recognition using desk and wearable computer based video," IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE IN- TELLIGENCE, vol. 20, no. 12, pp. 1371–1375, 1998.

[13] D. H. Hu, S. J. Pan, V. W. Zheng, N. N. Liu, and Q. Y. 0001, "Real world activity recognition with multiple goals." in UbiComp, ser. ACM International Conference Proceeding Series, vol. 344. ACM, 2008, pp. 30–39.

[14] L. Liao, D. Fox, and H. Kautz, "Location-based activity recognition," in In Advances in Neural Information Processing Systems (NIPS. MIT Press, 2005, pp. 787–794.

[15] A. Zinnen, U. Blanke, and B. Schiele, "An analysis of sensor-oriented vs. model-based activity recognition," in IEEE Int. Symp. on Wearable Computers (ISWC), September 2009.

[16] U. Blanke and B. Schiele, "Remember and transfer what you have learned-recognizing composite activities based on activity spotting," in Wearable Computers (ISWC), 2010 Int. Symp. on. IEEE, 2010, pp. 1–8.

[17] S. Agarwal, A. Joshi, T. Finin, Y. Yesha, and T. Ganous, "A pervasive computing system for the operating room of the future," Mob. Netw. Appl., vol. 12, no. 2-3, pp. 215–228, Mar. 2007.

[18] M. Cheng, M. KanaiPak, N. Kuwahara, H. Ozaku, K. Kogure, and J. Ota, "Dynamic schedulingbased inpatient nursing support: applicabil- ity evaluation by laboratory experiments," Journal International Journal of Autonomous and Adaptive Communications Systems, vol. 5, no. 1, pp. 39 – 57, 2012.

[19] F. Naya, R. Ohmura, M. Miyamae, H. Noma, K. Kogure, and M. Imai, "Wireless sensor network system for supporting nursing context- awareness," Journal International Journal of Autonomous and Adaptive Communications Systems, vol. 4, no. 4, pp. 361 – 382, 2011.

[20] M. Staeger, P. Lukowicz, and G. Troster, "Implementation and evalua- tion of a low-power sound-based user activity recognition system," in Wearable Computers, 2004. ISWC 2004. Eighth Int. Symp. on, vol. 1, Oct 2004, pp. 138–141.

[21]   B.Altakrouri,G.Kortuem,A.Gruenerbl,K.S.Kunze,andP.Lukowicz, "The benefit of activity recognition for mobile phone based nursing documentation: A wizard-of-oz study." in ISWC. IEEE, 2010, pp. 1–4.