Where am I: Recognizing On-body Positions of Wearable Sensors

Kai Kunze[1] Paul Lukowicz[1,2], Holger Junker[2], Gerhard Tröster[2]

[1]Institute for Computer Systems and Networks UMIT Hall i. Tirol, Austria
csn.umit.at
[2]Wearable Computing Lab ETH Zurich, Switzerland
www.wearable.ethz.ch

Abstract. The paper describes a method that allows us to derive the location of an acceleration sensor placed on the user's body solely based on the sensor's signal. The approach described here constitutes a first step in our work towards the use of sensors integrated in standard appliances and accessories carried by the user for complex context recognition. It is also motivated by the fact that device location is an important context (e.g. glasses being worn vs. glasses in a jacket pocket). Our method uses a (sensor) location and orientation invariant algorithm to identify time periods where the user is walking and then leverages the specific characteristics of walking motion to determine the location of the body-worn sensor.

In the paper we outline the relevance of sensor location recognition for appliance based context awareness and then describe the details of the method. Finally, we present the results of an experimental study with six subjects and 90 walking sections spread over several hours indicating that reliable recognition is feasible. The results are in the low nineties for frame by frame recognition and reach 100% for the more relevant event based case.

1 Introduction

A promising approach to context and activity recognition is the use of motion sensors (predominantly accelerometers) attached to different parts of the user's body. Various types of activities ranging from simple modes of locomotion analysis [14] to complex everyday [2] and assembly tasks[8] have been successfully recognized using such sensors. One thing that most of the work in this area has in common is that it relies on sensors being placed at specific locations on the body. Typically this includes the wrists, the arms, legs, hips, the chest and even the head. Once a subset of locations has been chosen, the system is trained on this specific subset and will not function properly if the sensors are placed at a different locations. This implies that the user either has to explicitly 'put on' the sensors each time he/she dresses up or the sensors have to be permanently integrated into the individual pieces of clothing.

T. Strang and C. Linnhoff-Popien (Eds.): LoCA 2005, LNCS 3479, pp. 264–275, 2005.

For wide spread and more near term applications of context recognition it would be more convenient to place sensors in appliances and accessories that most of us carry with us on a daily basis anyway. Such appliances include mobile phones, PDAs, key-chains, watches, hearing aids, headphones, badges and other smart cards and soon maybe even displays in glasses. Most of these devices are equipped with electronics and some already have sensors, including motion sensors (e.g Casio camera watch).

1.1 Context Recognition with Standard Appliance

While convenient for the user, the use of standard appliances and accessories for context recognition poses considerable difficulties. The main problem is that one can never be sure which devices the user carries with him and where on the body they are located. A mobile phone might be in a trousers pocket, in a holster on the hips, in a jacket or shirt pocket or in a backpack. Similar is true for a PDA, a wallet or a key-chain. Even such location specific devices as a watch, headphones or glasses might sometimes be carried in a pocket or a backpack.

As a consequence any system using standard appliances for context recognition must address the following issues:

1. Enough different body locations must be covered by sensor enabled appliances to provide sufficient information for the recognition task at hand.
2. The system is either able to deduce where on the body each device is located at any given moment or the recognition algorithm is location invariant.
3. The system must either be able to deduce the orientation of each device or the recognition algorithm is orientation invariant.
4. If the recognition algorithm is not location invariant, then it must be able to deal with different combinations of locations.

The work described in this paper constitutes the first step in our effort to facilitate the use of such appliances for complex context recognition. It focuses on the second point: the device location. We show that the most common human activity, namely walking, can be recognized in a location independent way. We then demonstrate how the information that the user is walking can be leveraged to determine device location.

The motivation for starting with device location is twofold. First we consider it to be the most critical issue. Experience shows that people usually have several accessories with them and mostly carry them at different locations. As an example in a typical scenario the user might carry a key-chain in his trousers pocket giving us the leg information, a watch on the wrist, a mobile phone in a holster on the hip and a smart card in a wallet in a jacket pocket. Thus while a systematic study will eventually be needed to determine the most common locations for different applications, at this stage we assume that there are relevant situations where point one is satisfied. Concerning device orientation, it has been shown by [10] and [5] that it can be derived from three axis acceleration sensors by looking at either time periods where the norm of the acceleration signal is 1 (no motion just gravity) or in an approximation by looking at the low pass filtered signal. Finally while there is considerable research potential in optimizing an adaptive classifier, simple solutions also exist proving that it is feasible for a system to deal with variable locations. In the worst case, a separate classifier could be trained for every relevant combination of locations. Since for the classifier it only matters that a sensor is at a certain location, not which device it is embedded in, the number of combinations is reasonable. In addition, only a limited number of all possible combinations will be relevant for a given recognition task.

The second motivation for looking at device location is the fact that location information itself is an interesting context. As an example, knowing if the glasses are worn or if they are in a pocket can be an important clue to the user's activity.

1.2 Related Work and Paper Contributions

The potential of body-worn sensors for context and activity recognition has been demonstrated in many scientific papers [3, 2, 11, 15, 14, 9, 16, 5, 6].

Irrespective of the signal processing schemes and recognition algorithms used, the variety of approaches presented differ in a number of ways. While some approaches rely on different types of sensors (e.g.[11, 15, 14, 5]), others solely use a single type of sensor such as accelerometers ([3, 2, 9, 16, 6]) for the recognition task. Furthermore, the approaches may differ with respect to the type of context targeted for classification ranging e.g. from the classification of walking behavior (e.g. [14]) to the recognition of complex everyday activities [2]. There also exist approaches targeting the same recognition task using the same sensors, but differ with respect to the placement of the sensors. One common example for this is the recognition of walking behavior (level walking, descending and ascending stairs) with accelerometers (e.g. [16, 12]). Despite the many differences, one thing that most of the approaches have in common is the fact that their recognition engines are unaware of the sensor locations and simply rely on individual sensors being placed at specific locations on the body. Consider the case, where e.g. two accelerometers A_1 and A_2 are used for a certain recognition task. The recognition system works properly when sensor A_1 is placed on the wrist, and sensor A_2 mounted on the thigh which is the configuration which the system has been trained for. If the two sensors would be exchanged after training, the system's recognition performance would without debt decrease, if it was not aware of this change. To allow recognition systems to adapt to such changes, it must be aware of the locations of the individual sensors on the body.

To our knowledge the method described in this paper is the first published attempt to facilitate this functionality. This paper describes the details of this novel method and presents an extensive experimental evaluation showing that it produces highly reliable results. The experimental evaluation elaborates not just the final results but also gives an insight into the effect of the different, individual optimizations steps contained in our method.

The work that comes closest to ours has been presented by Lester et. al [7]. They introduced a method to determine if two devices are carried by the same person, by analyzing walking data recorded by low-cost MEMS accelerometers using the coherence function, a measure of linear correlation in the frequency domain. In addition Gellersen et. al have shown how a group of devices can be 'coupled' by being shaken together. Furthermore there were a number of attempts detect whether an appliance such as a mobile phone is on a pocket in the hand or on the table [4].

2 Approach

2.1 General Considerations

Our approach is based on the obvious observation that different parts of the body tend to move in different ways. As an example, hand motions contain much more higher frequency components and larger amplitudes than hip or head motions. In addition, physiological constraints mean that certain types of motions are not permissible at all for some parts of the body (you can not turn your leg around the vertical axis in the knee or tilt your head more than 90 degrees). Also, some parts tend to be motionless for longer periods of time than the others. Thus, in theory, a statistical analysis of the motion patterns over a sufficient period of time should be able to provide information about the location of a sensor on the body. However, when implementing this idea in practice one has to deal with a number of issues. For one, the value of such a statistical analysis depends on the user activity during the analysis window. Little information will for example be gained if the user is sleeping during the whole time. In addition, the signal of a motion sensor placed on a given body part contains a superposition of the motion of this body part with the motion of the body as a whole. Thus while it is not possible to tilt the head more than 90 degrees, such a tilt will be registered when the user lies down. Finally, many of the motion characteristics that can be used to distinguish between body parts involve absolute orientation which is hard to detect, in particular if the orientation of the sensor is not known.

Our method deals with the above issues in two ways:

1. The analysis is constrained to the time during which the user is walking. This is motivated by two considerations. First, walking is a common activity that occurs fairly often in most settings. Thus, being able to detect the position of devices during walking phases should provide us with a sufficiently accurate overall picture of where the devices are located. In addition, once the location has been determined during a walking phase, this knowledge can be used to detect possible changes in placement.

   The second reason for focusing on walking is the fact that walking has such a distinct motion signature that it can be recognized without any assumptions about sensor location.
2. We base our analysis on the norm of the acceleration vector which is independent of the senor orientation.

2.2 Recognition Method

With the considerations described above our method can be summarized as follows.

Features Computation. Basic physical considerations confirmed by initial tests have lead us to use following features computed in a sliding window that is 1 sec long (overlapping 0.5 sec):

RMS: $\sqrt{\frac{1}{N} * \sum_i x_i^2}$, where N is the number of samples a sliding window contains, and x_i the i'th sample of the window.

75%Percentile: Given a signal s(t) the 75%percentile, also known as the third quartile, is the value that is greater than 75% percent of the values of s(t) .

InterQuartileRange: The interquartile range is defined as the difference between the 75th percentile and the 25th percentile.

Frequency Range Power: Computes the power of the discrete FFT components for a given frequency band.

Frequency Entropy: The frequency entropy is calculated according to the following formula: $H_{freq} = -\sum p(X_i) * log_2(p(X_i))$, where X_i are the frequency components of the windowed time-domain signal for a given frequency band and $p(X_i)$ the probability of X. Thus, the frequency entropy is the normalized information entropy of the discrete FFT component magnitudes of the windowed time-domain- signal and is a measure of the distribution of the frequency components in the frequency band. This feature has been used by [1].

SumsPowerWaveDetCoeff: describes the power of the detail signals at given levels that are derived from the discrete wavelet transformation of the windowed time-domain signal. This feature has successfully been used by [13] to classify walking patterns with acceleration sensors.

Training. Using a selection of relevant device positions walking recognition is trained. The recognition is trained in a location independent manner by putting the data from all locations into a single training set. As will be elaborated in the experiments section 3, best results were achieved with a C 4.5 classifier. However Naive Bayes, Naive Bayes Simple and Nearest Neighbor have also produced acceptable recognition rates. In the next phase, data collected during walking is used to train the location recognition.

Recognition. The recognition is performed separately by each sensor using the system trained according to the method described above. It consists of the following steps:

1. *Frame by Frame Walking Recognition* In this phase the features are computed in a sliding window of length 1s as described above and each window is classified as walking or non walking. The window length has been selected such that in a typical case it contains at least one step.

Fig. 1. Sensor placement

2. *Walking Recognition Smoothing* Using another jumping window of length 10 sec jumping by 5 sec the results of the frame by frame walking classification are then smoothed. The smoothing retains only those windows, where more then 70% of the frames were classified as walking. This ensures that the subsequent location classification is based only on 'clean' walking segments.
3. *Walking Segment Localization* The smoothed frame by frame recognition results are then used to localize walking segments that are long enough to allow reliable recognition. We define appropriate length to be at least a few tens of seconds and not longer than a 2 or 3 min. If a walking segment is longer than this boundary, it is automatically divided into several segments. The rationale behind this approach is that most devices are likely to remain in the same place for a few minutes. Changes on a smaller timescale must be considered as isolated events (e.g taking out a phone and rejecting an incoming call) and have to be detected separately by each device.
4. *Frame By Frame Location Recognition* A sliding window of the length of 1 sec. is then applied in each segment that has been identified as a relevant walking event. In each window the features for location recognition are computed and classification is performed.
5. *Event Based Location Recognition* For each segment a majority decision is performed on the frame by frame location classification.

3 Experimental Results

3.1 Experimental Setup

To evaluate the performance of our method an experiment was conducted with 6 subjects. For each subject, 3 experimental runs were recorded. Each run was between 12 and 15 min and consisted of the following set of activities:

1. Working on a desk (writing emails, surfing, browse through a book).
2. Walking along a corridor.

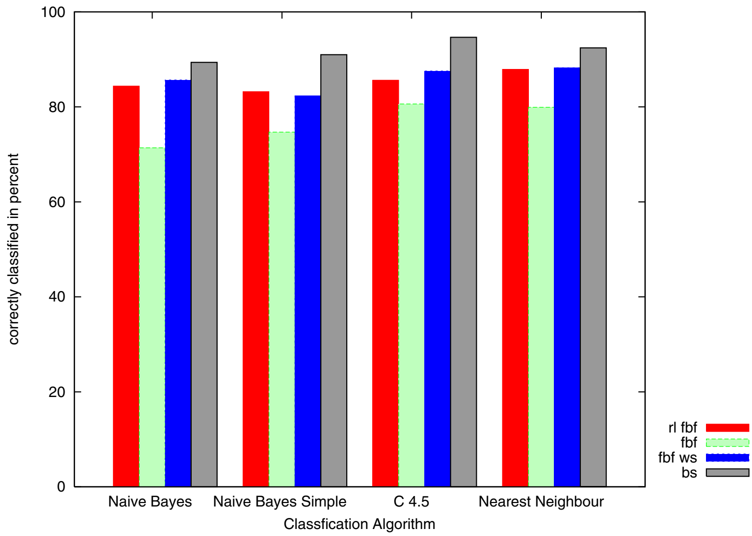


Fig. 2. Overview over the different classification algorithms and the varying approaches. The abbreviations have the following meaning: **rl fbf**=frame by frame using reference labeling (3.3), **fbf** = for frame by frame location recognition using frame by frame walking (3.3 and 2.2), **fbf ws** = frame by frame location recognition using smoothing over walking (2.2), **bs** = smoothing approach for both location and walking (2.2)

3. Making coffee, cleaning coffee kitchen, opening/closing drawers.
4. Walking along a corridor.
5. Giving a 'presentation'.
6. Walking.
7. Walking up and down a staircase.
8. (optional) working at desk.

Thus, there are 18 data sets containing a total of 90 walking segments (including the stairs).

All results presented below are based on a 10 fold-cross validation for evaluation on a subject by subject basis. For classification the WEKA [1] java package was used.

Hardware. The sensor system used for the experiment is the XBus Master System (XM-B) manufactured by XSens [2]. For communication with the XBus a Bluetooth module is used, thus data is transmitted wirelessly. Therefore, the test subjects just have to carry the XBus plus sensors and are not burdened with any additional load. The conductor of the experiment carries a Xybernaut to collect

[1] http://www.cs.waikato.ac.nz/ml/weka/
[2] http://www.xsens.com

the data and to supervise the experiment. Four of the XSens Motion Tracker sensors connected to the XBus are affixed on the test subject on four different body parts. The locations that have been chosen represent typical locations of appliances and accessories. Furthermore, they are relevant for the context recognition. Below is a short description of the locations used:

– Sensor 1: Wrist. This simulates a watch or a bracelet while worn.
– Sensor 2: Right side of the head above the eyes. This emulates glasses that are being worn.
– Sensor 3: Left trouser's pocket. This is a typical location for a variety of appliances such as key chains, mobile phones or even a watch that was taken off the wrist.
– Sensor 4: Left breast pocket. Again a typical location that would also include smart cards, glasses, (e.g. in a wallet).

3.2 Location Recognition on Segmented Data

As already mentioned earlier, the location recognition is only done during walking. Thus we begin our analysis by looking at the performance of the location recognition on hand picked walking segments. The results of the frame by frame recognition an all 90 segments contained in the experimental data is shown in figure 2. Using a majority decision on each segment leads to a 100% correct recognition (124 out of 124). The smallest segment is 1 minute long.

3.3 Continuous Location Recognition

Walking Recognition. The first step towards location recognition from a real life, continuous data stream is the detection of walking segments. As shown in Table 1 a frame by frame walking recognition (walking vs. not walking) showed an accuracy between 69% and 95% (mean 82%). However, for our purpose the mere accuracy is not the main concern. Instead we are interested in minimizing the number of false positives, as the subsequent location recognition works correctly only if applied to walking data. Here a mean of 18%(over all experiments) it is definitely to high.

As a consequence a false positive penalty has been added to the classification algorithms. Tests (see Figure 3) have lead to a minimal false positive rate considering a misclassification of 'Not Walking' four times worse than a misclassification of 'Walking'. While the overall correct rates goes down to between 61% and 85% (mean 76%), the percentage of false positives for 'Walking' is reduced to an average of 4% (between 0.5% and 7%).

The best results for the walking recognition is provided by the C4.5 tree algorithm with a mean of 82%, the worst by the Naive Bayes Simple with a mean of 65%.

In the next step the effect jumping window smoothing was investigated showing an average false positive rate of 2.17% with 84% of the windows being correctly recognized.

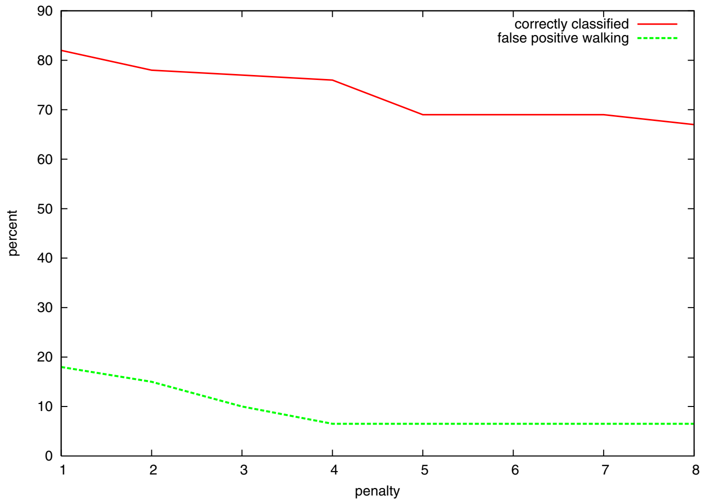


Fig. 3. Relation between correctly classified and false positives for walking

Table 1. Overview Classification for Walking in Percent

		Person1	P2	P3	P4	P5	P6	Mean
Frame by Frame	Correctly Classified	95	69	87	78	82	85	82.67
	False Positives for Walking	14	8	26	10	34	18	18.33
Frame by Frame with penalty	Correctly Classified	83	61	78	75	79	81	76.17
	False Positives for Walking	3	5	0.5	8	6	6	4.75
Frame by Frame penalty, jumping window	Correctly Classified	93	72	89	85	78	92	84,83
	False Positives for Walking	2	3	1	2	2	3	2.17

Walking Segments Location. In the last walking recognition step the walking segment location was applied to the smoothed frame by frame results. This has lead to 124 segments being located, none of which was located in a non-walking section. As shown for an example data set in figure 4 the only deviations from the ground truth was the splitting of single segments and the fact that the detected segments were in general shorter then the ground truth segments. However in terms of suitability for location recognition this is not relevant.

Frame By Frame Location Recognition. With the walking segments detected the frame by frame location recognition was applied. The results are shown in 2. They were later improved using the jumping window smoothing method which has lead to the results shown in 2 and 4.

The confusion matrices depicted in Table 3 indicate that the sensors attached to Head and Breast, as well as, Trousers and Wrist are most often confused. Especially, the confusion between Hand and Trousers is significant in size. One possible reason is that the movement pattern of Hand and Leg is similar while walking, particularly if the test subjects swings with the hand.

Table 2. Mean of C4.5 over all data sets for pre-labeled frame-by-frame (89,81 % correctly classified)

a	b	c	d	← classified as
856	2	87	5	a = Head
21	804	0	12	b = Trousers
101	32	765	4	c = Breast
0	103	5	819	d = Wrist

Table 3. Mean of C4.5 over all data sets for frame-by-frame using frame-by-frame walking recognition (80 % correctly classified)

a	b	c	d	← classified as
567	4	94	4	a = Head
3	431	3	178	b = Trousers
83	32	678	10	c= Breast
12	155	24	754	d =Wrist

Table 4. Mean of C4.5 over all data sets for both smoothed walking and location (94 % correctly classified)

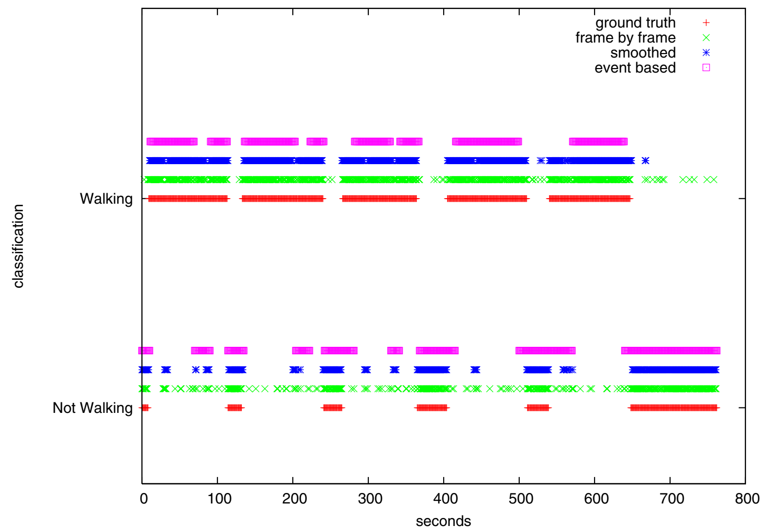
a	b	c	d	← classified as
965	2	31	2	a = Head
0	847	4	49	b = Trousers
42	0	883	1	c= Breast
17	68	10	921	d =Wrist

Event Based Location Recognition. In final step majority decision was performed in each segment leading to an event based recognition. Just like in the hand segmented case *the recognition rate was 100 %.*

4 Conclusion and Future Work

The work described in this paper constitutes a first step towards the use of sensors integrated in standard appliances and accessories carried by the user for complex context recognition. It is also motivated by the relevance of device location for general user context.

We have introduced a method that allows us to recognize where on the user's body an acceleration sensor is located. The experimental results presented above indicate that the method produces surprisingly reliable results. The method has found all walking segments in each experiment and has produced perfect event based recognition. Note that for practical use such event based recognition and not the less accurate frame by frame results that are relevant.



h

Fig. 4. Sample set containing different approaches for recognizing the walking segments

Despite this encouraging results it is clear that much work still remains to be done. The main issue that needs to be addressed is the detection of location changes that happen when the user is not walking and short duration location changes occurring during walking (e.g. taking out a mobile phone events during walking). Here, we see communication and cooperation between different devices as the key. Thus, for example, if all devices but the one located in a breast pocket detect walking then it must be assumed that something is happening to this device. Similar conclusions can be drawn if devices known to be in the trousers side pockets detect the user sitting and all but one devices report no or little motion while a single devices detects intensive movement. How such device cooperation can be used in real life situations and how reliable results it can produce is the subject of the next stage of our investigation.

Another interesting issue is the study of how well the recognition method can distinguish between locations that are close to each other on the same body segment. Finally we will need to see how well the actual context recognition works with standard appliances given only approximate location and not a sensor tightly fixed to a specific body part.

References

1. L. Bao. Physical activity recognition from acceleration data under semi-naturalistic conditions. Master's thesis, MIT, 2003.
2. L. Bao and S.S. Intille. Activity recognition from user-annotated acceleration data. In F. Mattern, editor, *Pervasive Computing*, 2004.

3. Ozan Cakmakci, Joelle Coutaz, Kristof Van Laerhoven, and Hans-Werner Gellersen. Context awareness in systems with limited resources.
4. Hans W. Gellersen, Albercht Schmidt, and Michael Beigl. Multi-sensor context-awareness in mobile devices and smart artifacts. *Mob. Netw. Appl.*, 7(5):341–351, 2002.
5. N. Kern, B. Schiele, H. Junker, P. Lukowicz, and G. Tr̈oster. Wearable sensing to annotate meeting recordings. In *Proceedings Sixth International Symposium on Wearable Computers ISWC 2002*, 2002.
6. N. Kern, B. Schiele, and A. Schmidt. Multi-sensor activity context detection for wearable computing. 2003. European Symposium on Ambient Intelligence.
7. J. Lester, B. Hannaford, and G. Borriello. "are you with me?" - using accelerometers to determine if two devi ces are carried by the same person. In A. Ferscha and F. Mattern, editors, *Pervasive Computing*, 2004.
8. P. Lukowicz, J. Ward, H. Junker, M. Staeger, G. Troester, A. Atrash, and S. Starner. Recognizing workshop activity using body worn microphones and accelerometers. In *Pervasive Computing*, 2004.
9. J. Mantyjarvi, J. Himberg, and T. Seppanen. Recognizing human motion with multiple acceleration sensors. In *2001 IEEE International Conference on Systems, Man and Cybernetics*, volume 3494, pages 747–752, 2001.
10. D. Mizell. Using gravity to estimate accelerometer orientation. In *Proceedings Seventh International Symposium on Wearable Computers IS WC 2002*, 2003.
11. A. Schmidt, K. A. Aidoo, A. Takaluoma, U. Tuomela, K. Van Laerhoven, and W. Van de Velde. Advanced interaction in context. *Lecture Notes in Computer Science*, 1707:89–93, 1999.
12. M. Sekine, T. Tamura, T. Fujimoto, and Y. Fukui. Classification of walking pattern using acceleration waveform in elderly people. In D. Enderle, J., editor, *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society Cat*, volume 2, 2000.
13. M. Sekine, T. Tamura, T. Fujimoto, and Y. Fukui. Classification of walking pattern using acceleration waveform in elderly people. *Engineering in Medicine and Biology Society*, 2:1356 – 1359, Jul 2000.
14. L. Seon-Woo and K. Mase. Recognition of walking behaviors for pedestrian navigation. In *Proceedings of the 2001 IEEE International Conference on Control Applications (CCA'01) (Cat*, pages 1152–1155, 2001.
15. K. Van-Laerhoven and O. Cakmakci. What shall we teach our pants? In *Digest of Papers. Fourth International Symposium on Wearable Computers.*, pages 77–83, 2000.
16. P. H. Veltink, H. B. J. Bussmann, W. de Vries, W. L. J. Martens, and R. C. Van-Lummel. Detection of static and dynamic activities using uniaxial accelerometers. *IEEE Transactions on Rehabilitation Engineering*, 4(4):375–385, Dec. 1996.