

Bayesian Recognition of Safety Relevant Motion Activities with Inertial Sensors and Barometer

Korbinian Frank, Estefania Munoz Diaz, Patrick Robertson
Institute of Communications and Navigation
German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
{firstname.lastname}@dlr.de

Francisco Javier Fuentes Sánchez
E.T.S.I. de Telecomunicación
Universidad de Málaga
Málaga, Spain
franciscojavier.fuentessanchez@gmail.com

Abstract—Activity recognition has been a hot topic in research throughout the last years. Walking, standing, sitting or lying have been detected with more or less confidence, in more or less suitable system designs. None of these systems however has entered daily life, neither in mass market, nor in professional environments. What is required is an unobtrusive system, requiring few resources and – most important – recognizing all important activities with high confidence. To this end, our research has focused on the professional market for safety related applications: first responders or also military use. Next to the classical motion related activities, our system supports motions in three dimensions that are necessary for all kinds of movements indoors as well as outdoors. These include falling, wriggling, crawling, climbing stairs up and down and using an elevator. We have proven this approach to run in real-time with only a single wireless sensor attached to the body while achieving robust and reliable recognition with a delay lower than two seconds.

Index Terms—Activity recognition; motions; crawling; 3D activities; Bayesian; Grid-based filter

I. INTRODUCTION

The recognition of activities like walking, standing, sitting, or lying has been a hot topic in recent years. The community has seen approaches with different kinds of sensors, first with many sensors spread around the body, later with a single sensor, but with a reduced amount of recognized activities. Although there are already stable prototypes, none of these systems has entered the daily life, neither in mass market nor in professional services. This is because these systems are obtrusive, not suitable for all situations or just too power-consuming.

Our work tries to overcome these problems. We have developed a system using sensors which are available in today's smart-phones: accelerometers, gyroscopes and barometers integrated in an inertial measurement unit (IMU) at one place of the body. We have added a wider set of activities like the usage of stairs, escalators and elevators, falling, running, jumping, wriggling or crawling. With this set of activities not only typical daily behavior can be recognized, but also professional users in safety relevant situations, like first responders or armed forces, can benefit from it. For these user groups, controlling agencies need reliable information about the team members in real-time to be able to quickly react on unforeseen events and emerging danger. Moreover the activity information

supports pedestrian dead reckoning systems used for first responders e.g. in indoor scenarios like malls or airports.

The objective of this work is to present and to evaluate the unobtrusive activity recognition system we have developed for safety critical operations. It recognizes the activities *sitting, standing, elevator down, elevator up, walking, walking down stairs, walking up stairs, running, jumping, falling, lying, low crawling, and high crawling* using Bayesian classifiers.

The remainder of this paper shall describe the complete development of the system. After an overview of related work in activity recognition (section II), the collected training data set along with the used sensor and our experiments are described in section III. The main part of this work has been the identification of a comprehensive set of significant features for the identified motion activities. Their nature and bio-mechanical background are presented in section IV. These features have been used in the classifiers, based on (supervised) learning techniques and expert knowledge. The resulting classifier is presented in section V before the evaluation results of the system under semi-naturalistic conditions are presented in section VI. Our conclusions and the outlook on the system with a displaceable sensor close this paper.

II. RELATED WORK

The related work on activity recognition is numerous and differs in many aspects. This starts with the activities under investigation, the sensors used as well as the features and classifiers. Among the activities studied in the literature are [1]:

- Human motion activities like *walking, walking upstairs and downstairs, standing or sitting*.
- Sports activities such as *jogging, cycling, rowing, callisthenics or martial arts moves*.
- Gestures such as *open door, close door or turning* mainly for industrial environments.
- *Activities of Daily Life (ADL)* such as *cooking, repairing a bike or entering a building*.
- The recognition of *emotions or stress* levels.

With respect to the sensors, many research works use external sensors, combine several sensors on different places of the body, or multimodal sensors on one part of the body. Thereby the sensor positions also vary greatly. Among the first

attempts were visual methods [2]. The major inconvenience is the previous installation of cameras and, in consequence, that the area to track the subject is limited. But still, video is used for instance in [3], [4] or [5].

A combination of several sensors at different parts of the body has been proposed among others by [6] (chest and thigh), [7] (chest and hip), [8] (left thigh, right ankle), [9] (wrist and thigh), [10] (right thigh, necklace, right and left wrists), [11] (torso, left and right sleeves, left upper and lower arm, left hand), and [9] (wrist and thigh).

More and more, inertial sensors [12] are used. Inertial sensors often combine accelerometers, gyroscopes and magnetometers in a multimodal sensors. They can also be combined with altimeters or GPS sensors and even integrated in cloths for the detection of breathing, speech or motions as in [13], [14] or [15]. More common however is the use of sensors embedded in mobile phones. Current models are able to provide information through GNSS, visual, auditive, light, temperature, direction, turn rate, pressure and acceleration sensors. Using a mass-market device, millions of people are able to take part in studies of daily habits [16]. Even moving sensor positions are therefore investigated, e.g. in [17]. Still, most research uses standalone inertial sensors in order to have better, drift-reduced measurements as long as the integrated sensors still do not meet the performance of standalone devices. They are worn on the foot [18], hip [19], chest [20], wrist [21] or leg [22]. Publications with inertial sensors of all kinds at only once position encompass [23], [23], [6], [10], [8], [21], [24], [11], [25], [26], [27], [22], [19], [20], [18], [7], [28], [29] and many others. The sampling rate thereby varies greatly, frequencies between 0.1 Hz and 200 Hz are common.

The same large variations exist with respect to the classification techniques and the features used for them. While many research groups use Bayesian techniques [18], also decision trees (e.g. [17], [15]), support vector machines (SVM) and principal component analysis (PCA, e.g. [30]), and many others are used. The features that discriminate the activities in the classifiers are sometimes differentiated in statistics and physical features [29]. The first type uses mathematical operators based on statistics, while the second one uses a physical interpretation of human motion. Often used features include variance (e.g. for walking, jogging, hopping), correlation between each pair of sensor axes, zero crossing rate, mean crossing rate, derivatives, mean, median, standard deviation, RMS, skewness, kurtosis, interquartile range, and spectral entropy (definition in a table), movement intensity (e.g. by a Euclidian norm), normalized Signal Magnitude Area (SMA), eigenvalues, averaged velocity, dominant frequencies, energy and many more. Also meta-research about features has been published. [15] compares the significance of sets of five features and even proposes a new feature, the *meantilt* and [30] learns the features for activity recognition and [20] publishes his *augmented feature vector*.

III. DATA SET

The first step in the development of this system was the creation of a representative data set of the respective activities from about 20 different persons under semi-naturalistic conditions, using an xsens MTw IMU. The test subjects (equally distributed among gender and age between early twenties and late fifties) have been given activity plans which they had to follow while they were observed and logged by a supervisor. All subjects were encouraged to perform the activities in various, individually characteristic ways. The data set has been labeled manually by the supervisor subsequently.

A. Sensor

The xsens MTw (see Figure 1) provides 3D angular velocity using rate gyroscopes, 3D acceleration using accelerometers, 3D earth magnetic field using magnetometers, as well as atmospheric pressure using a barometer [31] over a wireless connection. The battery is fully recharged after one hour and is told to last from 2.5 up to 3.5 hours in operation, in stand-by for approximately 90 hours. The physical characteristics are as follows [31]:

	Angular velocity	Acceleration	Magnetic field	Pressure
Dimensions	3 axes	3 axes	3 axes	-
Scale	$\pm 1200^\circ/s$	$\pm 160 m/s^2$	± 1.5 Gauss	0.3-1.1 Bar
Bias stability	$20^\circ/hr$	-	-	100 Pa/year
Noise/ \sqrt{Hz}	$0.05^\circ/s$	$0.003 m/s^2$	0.15 mGauss	0.85 Pa
Bandwidth	100 Hz (max)	100 Hz (max)	20 Hz (max)	-



Figure 1. Inertial measurement unit xsens MTw [31].

We have chosen the xsens MTw as it is small, light-weight, wireless and contains highly precise inertial sensors and a barometer. As it is wireless, however, we have to cope with the limited battery duration and the transmission distance of about 20 meters.

B. Experiment

In order to collect a valid data set, we have asked colleagues to execute five motion sequences. The test subjects wore the sensor fixed on their belts and have been followed by the experiment supervisor with a laptop receiving the sensor raw data. He guaranteed that the transmission distance is met and marked activity transitions by pressing a key. This recorded the timestamp of the received data. The activity sequences were the following:

- 1) Standing - Walking - Standing - Running - Standing
- 2) Standing - Waking up stairs - Standing - Walking down stairs - Standing

- 3) Standing - Elevator Up/Down - Standing
- 4) Standing - Jumping - Standing - [Falling - Lying] $\times 3$ - High Crawling - Standing - Sitting
- 5) Standing - Low Crawling - Standing

In our data set, we have tried to achieve a big number, but also variety of subjects (regarding age, sex, weight and height). We also tried to take care that very short-term activities like jumping or falling are repeated long enough to record a critical number of events.

C. Data Set

Eventually, the data set we used to train the classifier includes 2 hours and 37 minutes of raw measurements. 20 volunteers between 23 and 58 years have taken part, among them 6 women and 14 men. This data set is made publicly available shortly. Please contact the authors for details.

In an offline learning phase, these measurements have been complemented by the xsens MTx data set (without barometers) from [19] provided under http://www.dlr.de/kn/desktopdefault.aspx/tabid-8500/14564_read-36508/.

IV. FEATURES

In order to ease classification, not the raw data are directly used, but some specially significant characteristics of the data, so called *features* are used. In the following, we shall explain how we can compute features and which ones we have selected for our purposes.

A. Feature Design

We consider features as the result of a chain of mathematical operators. This allows us to define a systematic creation of features, which helps us to distinguish sets of points associated to different classes independently from the operator nature (physics, mathematics, bio-mechanics,...), mathematical background (statistics, probabilistic, based on frequency analysis,...) or its complexity. Thereby we have identified three types of operators (see also Figure IV-A):

- Specific operators: this type of operator is directly applied to raw data.
- Generic operators: these operators use a vector as input and generate another vector as output.
- Mixer operators: two or more vectors are used as inputs and generate one vector as output.

The features used for activity recognition usually represent some bio-mechanical characteristic as a starting point. The application of a specific operator gives meaning to the data and has some characteristics that are highlighted with generic operators. Finally, mixer operators take advantage of the mix of several bio-mechanical characteristics linked to an activity class.

Example for a specific operator is the norm. Its output estimates the size of the vector of the magnitude in which we apply the operator. An example for a specific operator that is interesting for further processing is the acceleration in the vertical axis.

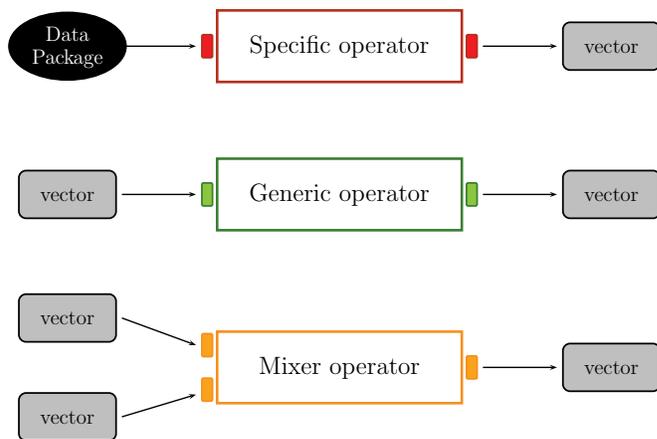


Figure 2. Types of operators used to create features from raw data.

The most important characteristic of generic operators is that they are completely independent of the meaning behind data. This way, the output of a generic operator can be concatenated with the input of any number of other ones. Its purpose can be detached from the physical meaning of the data, but only targeted to expose the distinction between data classes.

Sometimes several characteristics are very representative for one activity. In this case one could use single features for every characteristic and let the classifier make sense of them. In order to bring more expert knowledge to the classifier, however, we have defined a *mixer* operator (see third option in Figure IV-A) to take the information from different sources and to compress them into only one feature.

Next to the operators, also the frame of the raw data is important. There are the *sensor* (*SF*), *body* (*BF*) and *global* (*GF*) frames [12]. They can be generated from each other by *Euler angles*, *quaternions* or a *direction cosine matrix* (*DCM*). In particular the body frame, referring directly to the movements of the body, and the global frame, relating the sensor attached to the body to the earth (magnetic north and gravity direction), are of high importance for the remainder of this work.

As we have decided to use discrete Bayesian networks as classifiers, we also have the need to convert the continuous value ranges of the features to a discrete state space. In order to have more flexibility, we refrained from using clustering mechanisms, but discretized the value ranges with the help of histograms and feature plots in a way that allowed a close to optimal separation of activity classes.

B. Selected Features

In the search of the most adequate features, a couple of hundred possibilities have been tested and discarded. Eventually, we have selected 22 features with the highest impact (shown in Table I) for the activity classification.

In the following, the features will be described in groups according to their main purposes. In the supporting figures, the

No.	Definition	Window
1	$ a_{h_{BF}} $	128
2	$\sigma a_{h_{BF}} $	128
3	$MAX(a_{v_{BF}})$	128
4	$\overline{a_{v_{BF}}}$	128
5	$\sigma_{a_{v_{BF}}}$	128
6	$RMS_{a_{v_{BF}}}$	128
7	$IQR(\omega_{h_{BF}})$	128
8	$ a $	32
9	$\sigma a $	256
10	$\hat{E}(a _{BPF\ 1.6 - 4.5\ Hz})$	64
11	$MFC a $	128
12	$att a_{h_{BF}} , a_{v_{BF}}$	64
13	$RMS_{med}(a_{h_{GF}})$	100
14	$\sigma_{med}(\omega)$	100
15	$MIN(\overline{med(a_{h_{BF}})})$	100
16	$\rho_{a_{v_{BF}}, a }$	100
17	$\log_{10}(\overline{a_{v_{BF}}})$	25
18	$MAX(a_{h_{BF}})$	50
19	$MAX(2^{bollinger(\overline{med}(a_{v_{BF}} _{LPF[0.0\pi rad]})})$	181
20	$\hat{E}_{SD}(MoveToZero(\omega_{h_{GF}} _{LPF[0.0\pi rad]})$	208
21	$IUD_{150,150,0.8}(p)$	150
22	$\sum_{N=50} Polyfit(RMS(p))$	225

Table I

THE SET OF FEATURES USED FOR ACTIVITY RECOGNITION. *No.* SHOWS THE FEATURE NUMBER USED IN THE FOLLOWING, THE OVERALL WINDOW LENGTH *Window* IS DEFINED IN NUMBER OF SAMPLES (IN OUR CASE AT 100Hz).

relevant activity instances in the plots have been color-coded according to Table II:

1		Sitting	8		Running
2		Standing	9		Jumping
3		Elevator down	10		Falling
4		Elevator up	11		Lying
5		Walking	12		Low crawling
6		Walking upstairs	13		High crawling
7		Walking downstairs	14		Transition

Table II

CODING (COLOR OF INDEX) OF ACTIVITY CLASSES FOR THE REMAINDER OF THIS PAPER.

1) Distinction between Dynamic and Static Motion States:

Features 1, 2, 13, and 14 are particularly useful to distinguish between dynamic (all kinds of walking, running, jumping, falling and crawling) and static (sitting, standing, lying and elevator usage) activities.

Features 1 and 2 therefore use characteristics of horizontal acceleration in the body frame as these contain information about the body attitude. Feature 1, $|a_{h_{BF}}|$, is the mean value of the acceleration on the x and y axes in a window of 128 samples – a length which is relevant for both instantaneous and repeating activities.

Feature 2, $\sigma|a_{h_{BF}}|$, then represents the standard deviation of the same base signal over a window of 128 samples. This feature measures the variation over $|a_{horiz_{BF}}|$ for every activity. A window length of 128 samples is chosen in order to be a meaningful feature for all the activities considered.

Static from non-static activities can be distinguished except for *walking*.

Feature 13, $RMS_{med}(a_{h_{GF}})$, is the root mean square of the median of the acceleration in the horizontal plane of the global frame. The root mean square uses a window of 97 samples and the median 3 samples, Feature 14, $\sigma_{med}(|\omega|)$, computes the standard deviation (97 samples) of the median (3 samples) of the norm of the angular velocity of the user. These features indicate how strong or weak some motion is.

The acceleration in the horizontal plane for the global frame gives information about the displacement of the subject. For dynamic activities the values are higher than for static activities. The angular velocity vector complements this by information about the turn of the waist. Running or crawling have high rates of angular velocity as well as falling (within a very short time).

In order to get a clear picture, a plot with a classification of activities in two sets is drawn. Figure 3 shows that static activities (blue) have values around zero and the dynamic ones (red) move away from the origin.

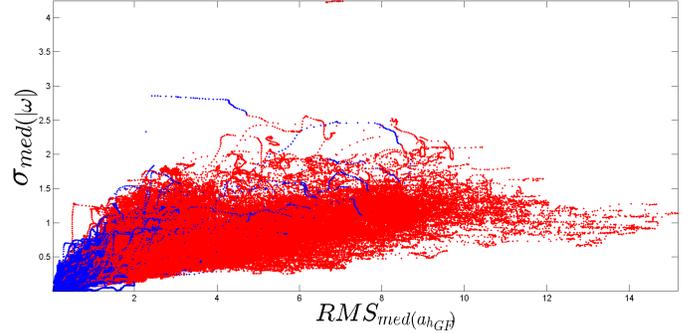


Figure 3. Activity plot for features 13 (x-Axis) and 14 (y-Axis) with the data set from section III divided into two groups, static (blue) and dynamic (red) activities.

2) Distinction among Static Activities:

One of the most difficult distinctions between our activities are *standing* vs. *sitting*. Therefore and for the further static activities, we have identified four particularly relevant features: Features 12, 15, 16, and 17. In addition, we assume that the system is started while the user is standing. Hence the position of the sensor during *standing* is known and the sensor measurements can be rotated to BF.

Feature 12, $att|a_{h_{BF}}|, a_{v_{BF}}$, gives information about the body attitude during the performance of the activity as it uses the interaction of the gravity field of the Earth in the horizontal plane and the vertical axis. It is defined in a 64 samples window size and calculated using the variation of the $|a_{h_{BF}}|$ and $|a_{v_{BF}}|$ between the current activity and the initial activity *standing*. However, we rely here on the assumption that the sensor is tilted why a person is sitting. It does not work with unusual sensor placements.

Feature 15, $MIN(\overline{med(|a_{h_{BF}}|)})$, is the minimum (within 47 samples) of the standard deviation (in 50 samples) of the median (among 3 samples) of the norm of the acceleration

in the horizontal plane of the body frame and Feature 16, $\rho_{a_{v_{BF}}, |a|}$, is the correlation over 100 samples of the vertical acceleration in the body frame and the norm of the acceleration (cf. Figure 4). Feature 17, $\log_{10}(\overline{a_{v_{BF}}})$, represents the decimal logarithm of the average acceleration in the vertical axis of the body frame over only 25 samples (cf. Figure 5).

Also these features reference the body frame and try to detect where the components of the gravity vector are incident. This detection is done analyzing the acceleration in the horizontal plane and the vertical axis. If the acceleration in the vertical axis is similar to the gravity value and almost zero in the horizontal plane, the body is upright. In contrast, if the acceleration in the vertical axis is close to zero and the horizontal plane contains the gravity, the body is in a horizontal position. Then the first case allows recognizing standing and elevator activities and the second one lying. Accelerations in both, horizontal plane and vertical axis, occur in the intermediate situations, i.e. it detects that the body is tilted, which is the case for *sitting*.

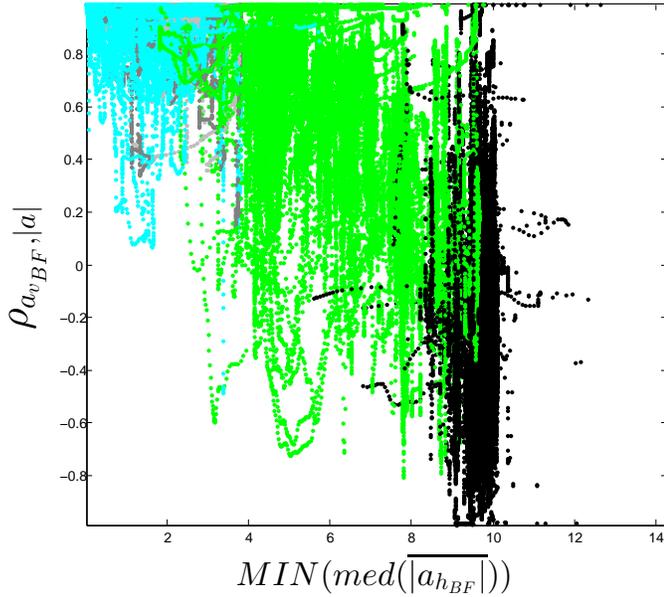


Figure 4. Activity plot for features 15 (x-Axis) and 16 (y-Axis) with the static activity instances from the data set described in section III, color-coded according to Table II.

3) Distinction between Activities with High Accelerations:

To distinguish between activities with particularly high or short accelerations, *running*, *jumping*, and *falling*, we have chosen in total eleven features

Feature 3, $MAX(a_{v_{BF}})$, is the maximum value of $a_{v_{BF}}$ calculated in a window size of 128 samples. This window length is a compromise for the distinction between *jumping* and *falling* (short-term) and *walking* (repeating patterns). Feature 4, $\overline{a_{v_{BF}}}$, is the averaged vertical acceleration over 128 samples. This feature is very useful for distinguishing *standing*, *sitting* and *lying*. As soon as the wearer of the sensor hits the floor, a peak of the $a_{v_{BF}}$ is measured. These peaks are higher for activities such as *jumping* and *running* than *walking*

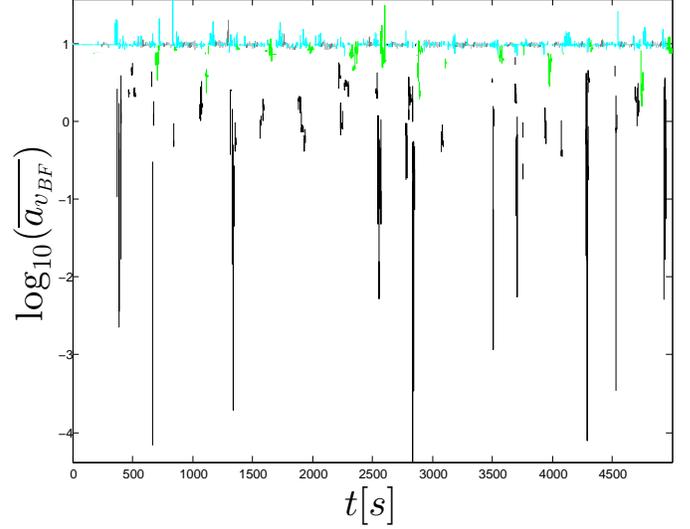


Figure 5. Activity plot for Feature 17 (y-Axis) over time (x-Axis) with the static activity instances from the data set described in section III, color-coded according to Table II.

and are the cause that the mean value of this signal increases for the former activities. Feature 5, $\sigma_{a_{v_{BF}}}$, is calculated as the standard deviation of $a_{v_{BF}}$, calculated in a window of 128 samples. This feature helps to build a good discriminator for the activities *jumping*, *running*, *walking* and *falling*. Feature 6, $RMS_{a_{v_{BF}}}$, is the root mean square of $a_{v_{BF}}$ for a window of 128 samples and also helps to distinguishing *lying* from *standing* and *sitting*

Feature 7, $IQR(|\omega_{h_{BF}}|)$, calculates the inter-quartile-range of $|\omega_{h_{BF}}|$ in a window of 128 samples and is particularly significant for *falling*. As *falling* implies a fast rotation of the body we have chosen this feature. A window of 128 samples is short enough to be strongly influenced by a fall.

The value of Feature 8, $|a|$, for a window length of only 32 samples is close to the current, instantaneous value of $|a|$. The oscillations in the value of $|a|$ imply that every activity can be observed. The main information given by this feature to the system is the characterization and distinction of the activity *jumping* from other activities. During the rising phase, $|a|$ decreases, in order to increase once the impulse has been taken. But after increasing the acceleration of the body, the acceleration value decreases considerably during the phase of free falling. On the other hand, after falling and hitting the floor, a big peak in the acceleration due to the reaction forces measured by the sensor is detected. Both values are very characteristic for *jumping*.

Feature 9, $\sigma_{|a|}$, is the standard deviation of $|a|$ in a 256 samples window. It is used to distinguish between *running* and the short term activities calculated in a window length of 256 samples.

Feature 10, $\hat{E}(|a|_{\text{BPF } 1.6 - 4.5 \text{ Hz}})$, calculates the energy of the filtered norm of the acceleration in a 64 sample window. The frequency content of measured acceleration during human motion related activities is below 10Hz [32]. The acceleration

norm measured during the performance of the activity *walking* has most of its energy in its first seven harmonics. Taking into account a fundamental frequency around 1.6 to 2 Hz, the energy of the signal will remain below 10 Hz. For *running*, the fundamental frequency is higher, around 2.5 Hz to over 3 Hz. Feature 11, $MFC_{|a|}$, is defined as the main frequency component of $|a|$, $MFC_{|a|}$, computed over a window size of 128 samples, in order to find a trade off between step periods while *running* and short activities like *jumping* and *falling*.

Finally, Feature 18, $MAX(|a_{h_{BF}}|)$, is the maximum normed acceleration in the horizontal plane of the body frame (calculated over 50 samples) and Feature 19, $MAX(2^{bollinger}(med(a_{v_{BF}})|_{LPF[0.0\pi rad]}))$, is calculated to extract the trend of the vertical acceleration in the body frame (181 samples in total). Both features are shown in Figure 6. We use the acceleration in the horizontal plane of the body frame and stress it by the maximum operator to detect the strength of the projection of the user in the lateral motion. It is highest in some moments of falling. The use of the acceleration in the vertical axis allows extracting the high values of the acceleration in the z -axis while jumping.

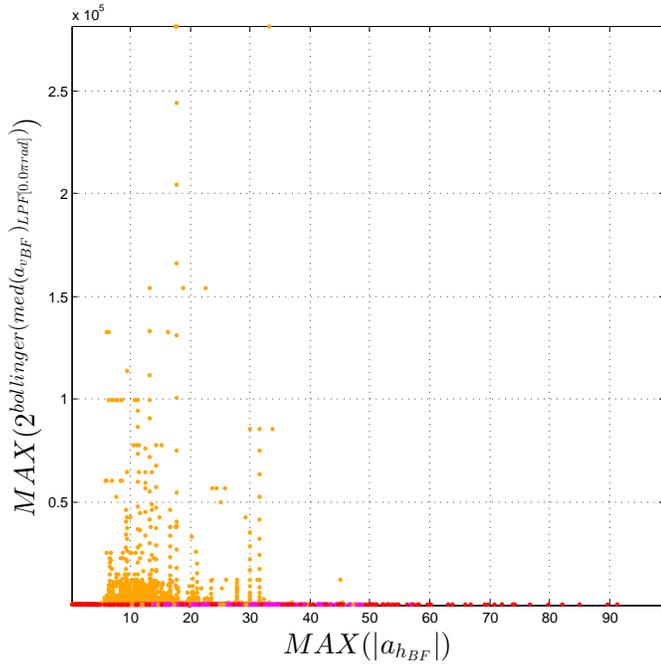


Figure 6. Activity plot for features 18 (x -Axis) and 19 (y -Axis) with the activity instances for running, jumping and falling from the data set described in section III, color-coded according to Table II.

4) *Distinction between Crawling States:* Feature 20, $\hat{E}_{SD}(MoveToZero(\omega_{h_{GF}}|_{LPF[0.0\pi rad]}))$, calculates the energy spectral density in frequency domain from a low-pass-filtered angular velocity in the horizontal plane of the global frame after removing the trend with the *MoveToZero* operator. It has been chosen to distinguish between high and low crawling, as it measures the turn of the hip during approximately two seconds (208 samples).

While standing we have the knee and the ankle between the

hip and the floor and we can regulate the space between both easily. Crawling has the characteristic that our hip is in contact with the floor without intermediate joints. Looking at human crawling (also in a resting and stable state), we appreciate that the hip is in contact with the ground through a fixed length. In the case of high crawling the length is the length of the femur and for low crawling the hip is directly on the ground. To move forward, one has to create space to move the lower limbs. In activities such as walking or running, the hip is slightly moved to create enough space, while the knee and the ankle help to reduce the turn of the hip and to accelerate the displacement. In contrast to walking or running, while crawling the lower limb cannot be folded in order to move which is why the hip has to be moved more. The less space the subject has, the bigger has to be the motion of the hip, i.e. the turn is highest for low crawling. The mathematical operators of Feature 20 highlight this characteristic (cf. Figure 7).

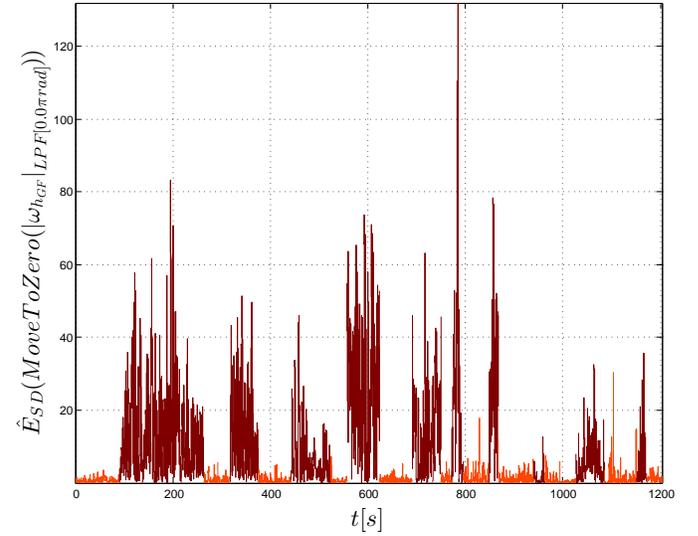


Figure 7. Activity plot for Feature 20 (y -Axis) over time (x -Axis) with the *crawling* instances from the data set described in section III, color-coded according to Table II.

5) *Distinction between 3-dimensional Motions:*

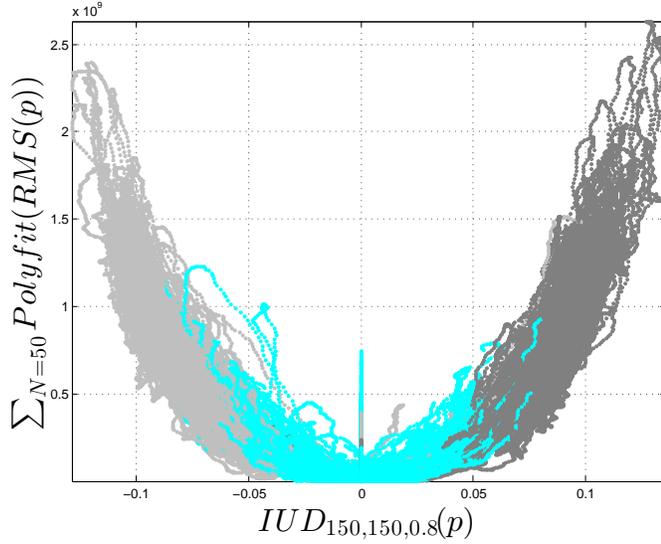
For two activity sets, walking and standing, there are variants related to motion up and down stairs or elevators. Feature 21, $IUD_{150,150,0.8}(p)$, and Feature 22, $\sum_{N=50} Polyfit(RMS(p))$, are meant to extract the trend of the atmospheric pressure p . Feature 21 uses a modified indicator up-down (IUD) with window length 150 for mean and update calculations and 0.8 for the environment around the standard deviation, Feature 22 is calculated as a sequence of root mean square, polyfit and sum over 225 samples.

An example of these features with our data set is given in Figure 8.

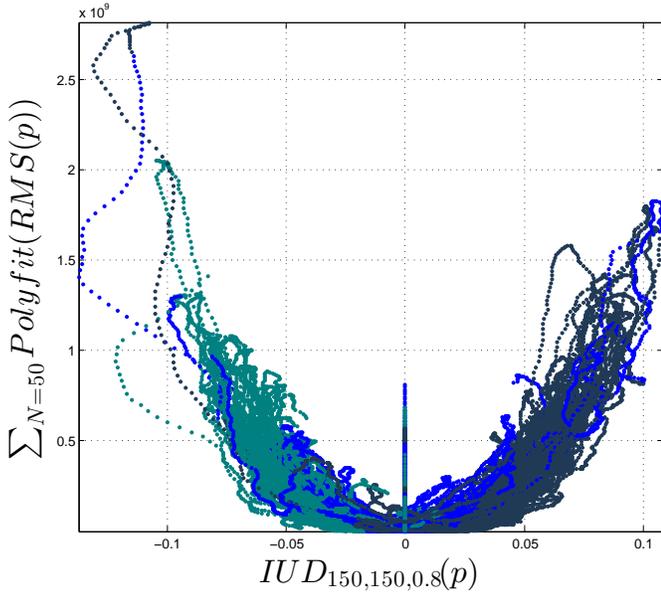
If a subject stands in a straight vertical position with almost no motion in any direction, still the altitude of the subject can change due to an elevator. In physical measurements it is detected by the variation of the atmospheric pressure, as acceleration and angular velocity do not change. The only

exception is an increase and decrease of the acceleration for the initial and final part of the motion of the elevator.

For walking and the usage of stairs, in contrast, also the accelerations and turn rates change. Still, the variation of the pressure is a good indicator.



(a)



(b)

Figure 8. Activity plot for features 21 (x -Axis) and 22 (y -Axis) according to the activity colors from Table II: (a) Activities *standing*, *elevator up* and *elevator down*; (b) Activities *walking*, *walking upstairs* and *walking downstairs*.

V. CLASSIFIER

The selected features shown in the previous section are the decision criteria for our classifier. Among numerous classification methods, we have chosen to use Bayesian techniques, as they can cope with uncertainty, erroneous or missing sensor

measurements. As their output is “soft” the result bears more information than a pure class.

Also within Bayesian techniques, there are many options. Basically, one can choose between a Naïve Bayes approach and a full Bayesian approach, as well as between static and dynamic information [33]. We have decided against a Naïve Bayes approach, as in our case always measurements from all sensors are available and therefore we do not expect significant performance advantages for the Naïve Bayes approach. Learning algorithms moreover show clearly better data-fitting for the full Bayesian approach. Hence we have employed a static and a dynamic Bayesian approach to classify activities. The corresponding Bayesian network and the grid-based filter shall be explained in the following sections.

A. Bayesian Network

A *Bayesian network (BN)* is a probabilistic model consisting of a Triplet (V, E, P) , with a set of *Random Variables (RVs)* $V = \{A_1, A_2, \dots, A_n\}$, a set of dependencies $E = \{(A_i, A_j) | i \neq j, A_i, A_j \in V\}$ between these RVs and a *joint probability distribution (JPD)* $P(V) = P(A_1 \cap A_2 \cap \dots \cap A_n)$. P is the product of the *Conditional Probability Distribution (CPD)* of every RV $P(A_i) \forall A_i \in V$. A BN must not contain directed cycles. They have been described among others by Pearl [34] or Jensen [35]. There are many approaches to “learn” Bayesian networks, i.e. the training of the classifier from data. One of the most used ones is the K2 algorithm introduced by Cooper and Herskovits in [36]. There are many implementations of it available online. We have used both a public implementation and for validation also our own version based on a *greedy hill climber* approach. Using a log-score metric we have chosen the best training result after 20000 iterations allowing a maximum of 6 parent nodes per random variable. This learning process took about 15 days and yielded the BN shown in Figure 9.

Although there are three feature nodes without direct connection to the RV *activity* representing the 14 desired states, f_{02} , f_{06} , and f_{19} , these cannot be omitted. The evidence represented by these RVs is not *d-separated* from *activity*, as their are active paths between via f_{09} , f_{03} , and f_{08} respectively.

B. Grid-based Filter

A dynamic Bayesian network has the advantage that the influence from the last activities can also be taken into account. In particular for motions this is valuable information. A first-order *Hidden Markov Model (HMM)* is the simplest case of a dynamic Bayesian network, but sufficient to be general. It can be shown that every temporal dependency can also be represented by a first-order HMM.

This first-order HMM for the activity recognition can be characterized by $\lambda \sim (A, B, \pi)$, where A is the transition matrix, matrix B represents the observations and π is the prior distribution of the initial states. Applied to the BN from the previous section V-A, the resulting HMM is shown in Figure 10.

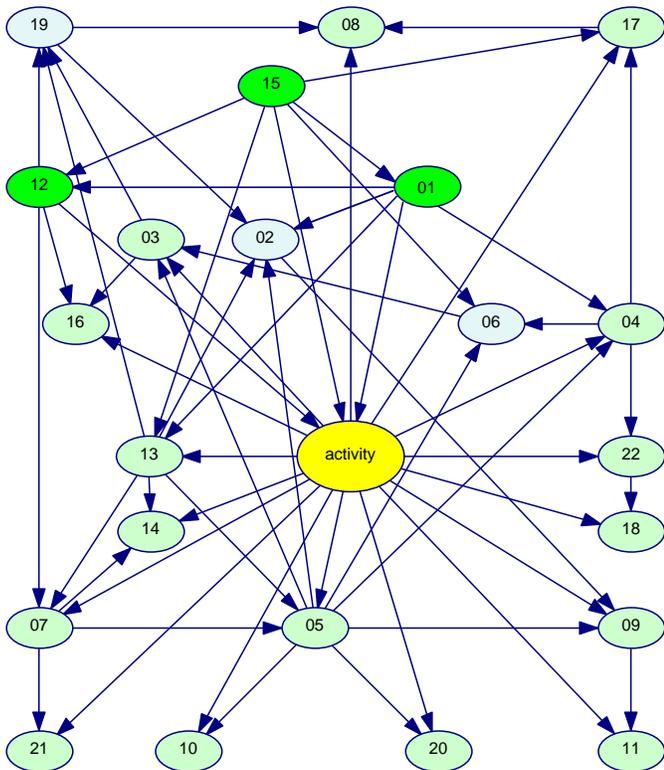


Figure 9. Bayesian Network classifier. The ellipsoids represent random variables. Their number is the identifier for the feature, the random variable *activity* has 14 states corresponding to the desired activities and a transition state. Parents of *activity* are colored green, children in light green.

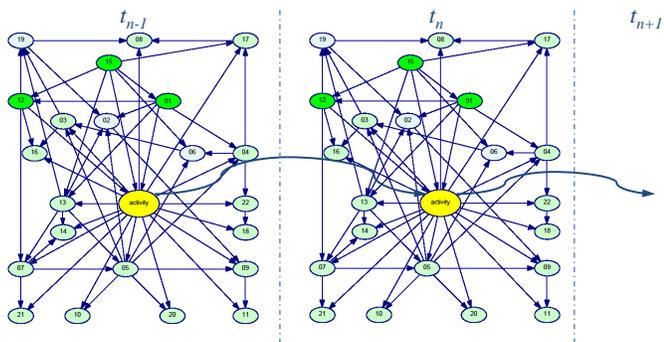


Figure 10. First-order Hidden Markov Model for activity recognition based on a learned Bayesian network.

Classification with such a HMM is the estimation of the most probable hidden state at time t given the past and current observations $O(1 : t) = O(1), O(2), \dots, O(t)$ as well as the model λ , $arg_i \max(p(a_i(t)|O(1), O(2), \dots, O(t), \lambda))$. As in our system the hidden state space has a finite number of states (i.e. activities), a grid-based filter [37] can be applied providing an optimal estimation of the posterior probability density function $p(a(t)|O(1 : t), \lambda)$. Once the posterior probability is estimated, the most probable activity is given by the state with the maximum probability.

We have created the HMM in a greedy process. We have

started without the option for transitions, i.e. with a diagonal matrix A . We improved this according to bio-mechanical expert knowledge, e.g. adding the transition *standing* to *sitting* at first. In a rapid-prototyping approach with short feedback loops, we added the allowed transitions. A first approach based on statistics of the data set did not show satisfactory results. The resulting transition matrix $A = (a_{ij})$, where $1 \leq i, j \leq 14$ and a_{ij} is the transition probability from a_i to a_j , is shown in Equation 1 (probabilities are rounded). The meaning of the indices follows Table II.

VI. EVALUATION

We have evaluated both classifiers developed. A first evaluation can be given based on the intrinsic metrics of the Bayesian learning algorithms and be based on the training set. This evaluation shall be complemented by an experiment conducted with our complete Java based implementation, i.e. taking into account delays and unknown data. Core measures for all evaluations (per activity class) are precision, recall and the f-measure:

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}$$

$$F-Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall},$$

where TP is the number of *true positives*, i.e. the instances correctly classified into a certain activity class, FP is the number of *false positives*, i.e. the instances incorrectly classified into a certain activity class, and FN is the number of *false negatives*, i.e. the not recognized instances that actually belong into a certain activity class.

A. Theoretical Evaluation of Classification with the BN

The theoretical evaluation refers to the static Bayesian network model shown in Figure 9. Its evaluation is based on the data set from section III, containing 32033 data instances. The evaluation on the training set yielded an overall percentage of 84.8% correctly classified instances.

The exact results per activity class are shown in Table III. Generally, precision and recall values between 85% and 95% are achieved, but also worse values for *falling* (around 75%), walking up and downstairs, as well as the precision of walking. These are explained by high confusions between these three activity classes stemming from the similar bio-mechanical background, to be compared with the confusion matrix from Table IV.

B. Naturalistic Real-time Evaluation

In order to evaluate the complete Java-based system, we conducted experiments with colleagues who had not participated in recording the data set. Hence the system has not been trained on their particular movements, body shapes, etc. We have recorded the following sequence of actions:

Standing–Sitting–Walking–Walking Upstairs – Standing – Walking Downstairs – Walking – Standing – Running – Standing – Walking – Elevator Up – Standing – Elevator Down –

$$A = \begin{pmatrix} .9463 & .0526 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .00001 & 0 & .0001 & .0011 \\ .0317 & .9497 & .0001 & .0001 & .0106 & .0001 & .0001 & .0053 & .0026 & .00001 & 0 & 0 & 0 & 0 \\ 0 & .3333 & .6667 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .3333 & 0 & .6667 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .0653 & 0 & 0 & .9326 & .0003 & .0005 & .0009 & .0005 & .00001 & 0 & 0 & 0 & 0 \\ 0 & .0099 & 0 & 0 & .0010 & .9881 & 0 & .0010 & 0 & .00001 & 0 & 0 & 0 & 0 \\ 0 & .0099 & 0 & 0 & .0010 & 0 & .9881 & .0010 & 0 & .00001 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .0099 & 0 & 0 & .9891 & .0010 & .00001 & 0 & 0 & 0 & .00001 \\ 0 & .0907 & 0 & 0 & .0009 & 0 & 0 & .0009 & .9074 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .9901 & .0099 & 0 & .00001 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .9899 & .0001 & .0001 & .0099 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .00001 & .9990 & .00001 & .0010 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .00001 & .9990 & .0010 \\ 0 & .5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .5 \end{pmatrix}$$

	Sitting	Standing	Elevator Down	Elevator Up	Walking	Walking Down	Walking Up	Running	Jumping	Falling	Lying	Crawling Low	Crawling High	Getting Up/Down
Sitting	1941	63	14	0	0	0	0	0	0	1	52	2	25	64
Standing	36	8082	111	105	63	18	47	2	9	3	1	1	8	260
Elevator Down	0	279	1710	5	0	0	0	0	0	1	0	0	0	13
Elevator Up	0	332	26	1840	0	0	0	0	0	0	0	0	0	6
Walking	0	74	0	1	2241	82	83	6	0	2	0	1	5	87
Walking Downstairs	0	21	0	0	132	866	21	39	18	8	0	0	13	35
Walking Upstairs	0	43	1	0	315	13	914	51	7	3	0	5	12	57
Running	0	8	2	0	11	25	28	1360	25	12	0	0	3	20
Jumping	0	5	1	0	5	5	22	50	1879	8	0	0	0	16
Falling	7	31	3	0	4	4	0	2	1	273	3	3	9	24
Lying	28	20	1	0	0	0	0	0	0	8	1389	23	1	34
Low Crawling	4	2	0	0	2	1	3	0	0	1	19	1634	58	98
High Crawling	21	9	0	0	40	6	15	0	0	4	1	28	1898	104
Getting Up/Down	108	485	11	10	103	63	102	16	28	46	87	123	132	1142

Table IV

CONFUSION MATRIX FOR THE LEARNED BAYESIAN NETWORK, CF. SECTION III. INSTANCES OF THE ACTIVITIES IN THE ROWS HAVE BEEN EVALUATED AS ACTIVITIES IN THE COLUMNS.

	Sitting	Standing	Elevator Down	Elevator Up	Walking	Walking Down	Walking Upstairs	Running	Jumping	Falling	Lying	Crawling Low	Crawling High
P	90.5	85.5	91.0	93.8	76.9	80.0	74.0	89.1	95.5	73.8	89.5	89.8	87.7
R	89.8	92.4	85.2	83.5	86.8	75.1	64.3	91.0	94.4	75.0	92.4	89.7	89.3
F	90.1	88.8	88.0	88.4	81.5	77.5	68.8	90.1	94.9	74.4	90.9	89.7	88.5

Table III

EVALUATION OF THE LEARNED BAYESIAN NETWORK. THE ACTIVITY CLASS *transition* IS LEFT OUT AS IT IS ONLY AN AUXILIARY STEP FACILITATING THE USAGE OF THE DYNAMIC BAYESIAN MODEL. **P** STANDS FOR *precision*, **R** REPRESENTS THE *recall* AND **F** THE *f-measure*.

Walking – Standing – Falling – Lying – Standing – Jumping – Standing – High Crawling – Standing – Low Crawling

We have followed the test subject with a computer recording data and the activity transition timestamps manually. Therefore there will be some measuring inaccuracies and inconsistencies

as it was difficult to minute the exact flow of activities (e.g. waiting for the elevator and stepping in, going down to the floor, standing up, etc.). The recorded data with the inferred activities have been evaluated together with the manually inserted groundtruth once for the Bayesian network and once for the classifier using the grid-based filter.

Figure 11 shows the evaluation of the static classifier where every 450ms the features have been computed and an inference process has been started. We can observe that there is some delay in the recognition and a high oscillation of the probability of the most probable activity class. We can see an unexpected confusion of *running* with *walking downstairs*. The fall has been missed and the following (short) lying period has been confused with *sitting*, just as a later *transition* period. The most certain recognition has been for *sitting* and both *elevator* directions. In total, an average (unweighted) precision of 51% is reached, and an average (unweighted) recall of 46%. These low values are certainly also due to the delay

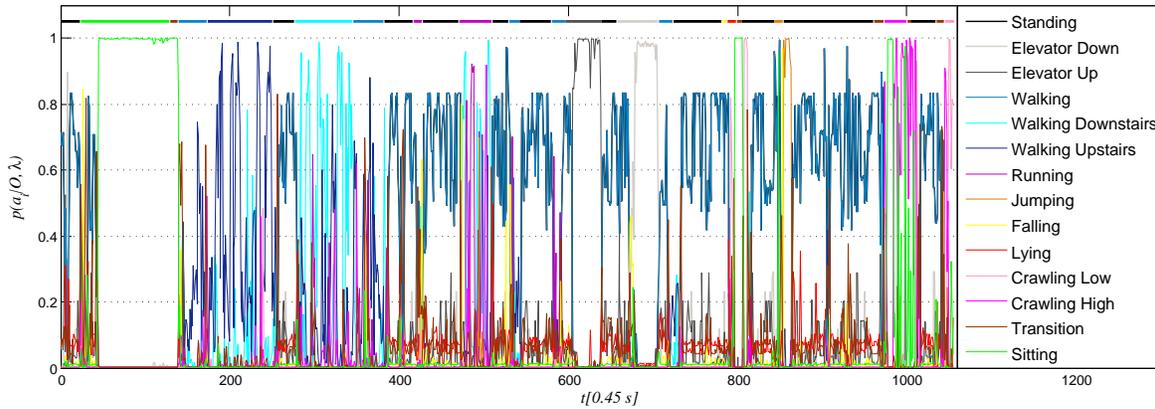


Figure 11. Evaluation of the static classifier with one test subject for about eight minutes. The top line shows the groundtruth, below we can see the inferred probabilities for every activity class according to the colors given in the legend on the right. Every data point corresponds to 450ms.

of the evaluation which is caused by the window length of the features and the processing time.

Figure 12 finally shows the evaluation of the same test subject’s activity sequence with the grid-based filter under the same evaluation conditions as above (same feature window lengths, evaluation for every 450 samples at 100Hz). We notice a much more stable behavior of the estimator, but a delay which is even a bit longer than for the BN. As above, it stems from window length and evaluation, but here also from the inertia of the HMM which always assigns higher probability to the previous activity. We see the same confusion between *running* and *walking downstairs* as with the BN classifier. The fall is misclassified as a transition, which must be because the test subject fell very slowly. Here, an average (unweighted) precision of 59% is reached, but an average (unweighted) recall of only 44%, although the plot shows that the estimator recognizes almost everything with high confidence. Again, this is due to the delay, which is particular severe for short-time activities.

VII. CONCLUSION AND OUTLOOK

Hence, with the approach suggested in this work, we present the first system which is resource-efficient, unobtrusive and reliable even for motions like wriggling, crawling, walking stairs and using escalators and elevators. It opens up new exploitation possibilities in the professional and the consumer market sectors.

The pending extension we are currently working on is the implementation of this system for Android operated smartphones using the internal sensors instead of the xsens IMU. The built-in low-cost sensors are expected to experience significantly more noise, which may entail modifications to the data set collection and feature discretization process. To be fully operational on an everyday basis, our system will be extended to recognize sensor posture, such as “pocket-based”, “hand held” and “swinging”. Using a Bayeslet approach, the most suitable recognition network shall be selected then. Combined with a usage based sensor management system, power consumption can be reduced to a minimum.

ACKNOWLEDGEMENTS

This work has been funded under grant agreement 257493 of the SOCIETIES Collaborative Project by the European Community’s 7th Framework Programme [FP7/2007-2013].

REFERENCES

- [1] D. T. G. Huynh, “Human Activity Recognition with Wearable Sensors,” Ph.D. dissertation, Technische Universität Darmstadt, August 2008.
- [2] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, “Hierarchical model-based motion estimation,” in *Computer Vision ECCV’92*. Springer, 1992, pp. 237–252.
- [3] M. Losch, S. Schmidt-Rohr, S. Knoop, S. Vacek, and R. Dillmann, “Feature set selection and optimal classifier for human activity recognition,” in *Proceedings of the 16th IEEE International Symposium on Robot and Human interactive Communication, RO-MAN 2007*. IEEE Computer Society Press, Aug. 2007, pp. 1022–1027.
- [4] M. S. Del Rose and C. C. Wagner, “Survey on classifying human actions through visual sensors,” *Artificial Intelligence Review*, vol. 37, no. 4, pp. 301–311, 2012.
- [5] B. Ni, G. Wang, and P. Moulin, “Rgbd-hudaact: A color-depth video database for human daily activity recognition,” in *Consumer Depth Cameras for Computer Vision*. Springer, 2013, pp. 193–208.
- [6] K. Aminian, P. Robert, E. Buchser, B. Rutschmann, D. Hayoz, and M. Depairon, “Physical activity monitoring based on accelerometry: validation and comparison with video observation,” *Medical & Biological Engineering & Computing*, vol. 37, no. 3, pp. 304–308, 1999, springer Verlag, ISSN: 0140-0118.
- [7] K. Frank, P. Robertson, M. Gross, and K. Wiesner, “Sensor-based identification of human stress levels,” in *MUCS 2013: 10th IEEE Workshop on Managing Ubiquitous Communications and Services part of PerCom 2013*, U. Hengartner and G. Schiele, Eds. IEEE, März 2013, pp. 127–132. [Online]. Available: <http://elib.dlr.de/80709/>
- [8] L. Bao and S. S. Intille, “Activity recognition from user-annotated acceleration data,” in *Proceedings of the Second International Conference on Pervasive Computing, PERVASIVE 2004*. Springer Verlag, Apr. 2004, pp. 1–17. [Online]. Available: <http://dx.doi.org/10.1007/b96922>
- [9] M. Uiterwaal, E. Glerum, H. Busser, and R. van Lummel, “Ambulatory monitoring of physical activity in working situations, a validation study,” *Journal of Medical Engineering & Technology*, vol. 22, no. 4, pp. 168–172, Jul. 1998, taylor & Francis, ISSN: 0309-1902.
- [10] S. Pirttikangas, K. Fujinami, and T. Nakajima, “Feature selection and activity recognition from wearable sensors,” in *Proceedings of the Third International Symposium on Ubiquitous Computing Systems, UCS 2006*, ser. Lecture Notes in Computer Science, H. Y. Youn, M. Kim, and H. Morikawa, Eds., vol. 4239. Springer Verlag, Oct. 2006, pp. 516–527.
- [11] T. Stiefmeier, “Real-time spotting of human activities in industrial environments,” Ph.D. dissertation, TU Darmstadt, 2008.

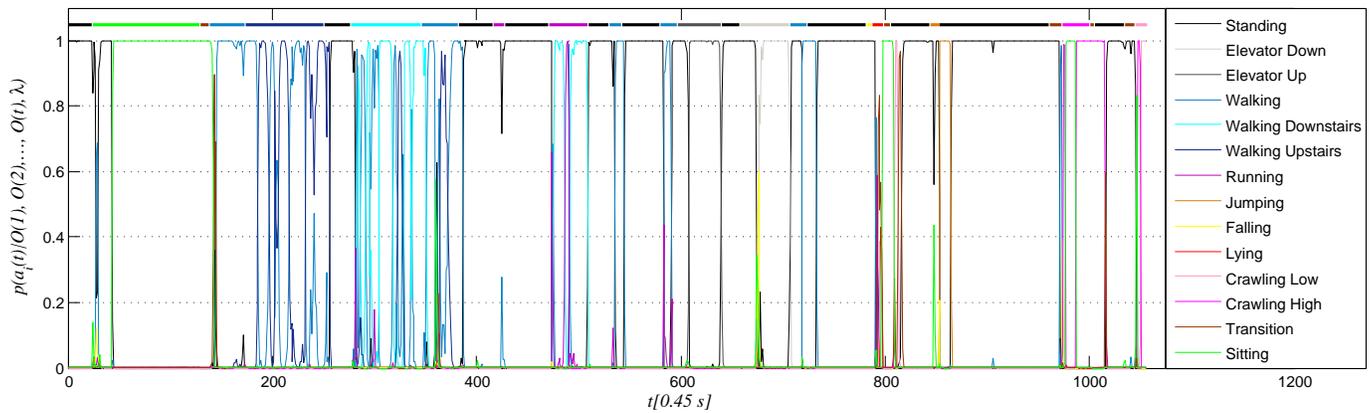


Figure 12. Evaluation of the dynamic classifier with one test subject for about eight minutes (same data as in Figure 11). The top line shows the groundtruth, below we can see the inferred probabilities for every activity class according to the colors given in the legend. Every data point corresponds to 450ms.

- [12] O. J. Woodman, "An introduction to inertial navigation," University of Cambridge Computer Laboratory, Tech. Rep., August 2007.
- [13] T. Yamada, Y. Hayamizu, Y. Yamamoto, Y. Yomogida, A. Izadi-Najafabadi, D. N. Futaba, and K. Hata, "A stretchable carbon nanotube strain sensor for human-motion detection," *Nature Nanotechnology*, vol. 6, no. 5, pp. 296–301, 2011.
- [14] L. Atallah, B. Lo, R. King, and G.-Z. Yang, "Sensor placement for activity detection using wearable accelerometers," in *Body Sensor Networks (BSN), 2010 International Conference on*. IEEE, 2010, pp. 24–29.
- [15] S. A. Muhammad, B. N. Klein, K. Van Laerhoven, and K. David, "A feature set evaluation for activity recognition with body-worn inertial sensors," in *Constructing Ambient Intelligence*. Springer, 2012, pp. 101–109.
- [16] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *SIGKDD Explor. Newsl.*, vol. 12, no. 2, pp. 74–82, 2011.
- [17] M. Susi, V. Renaudin, and G. Lachapelle, "Motion mode recognition and step detection algorithms for mobile phone users," *Sensors*, vol. 13, no. 2, pp. 1539–1562, Jan. 2013. [Online]. Available: <http://www.mdpi.com/1424-8220/13/2/1539>
- [18] J. Bancroft, D. Garrett, and G. Lachapelle, "Activity and environment classification using foot mounted navigation sensors," in *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, Nov. 2012, pp. 1–10.
- [19] K. Frank, M. J. Vera Nadales, P. Robertson, and M. Angermann, "Reliable real-time recognition of motion related human activities using mems inertial sensors," in *Proceedings of the 23rd International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS 2010)*, Sep. 2010, pp. 2919–2932. [Online]. Available: <http://elib.dlr.de/64996/>
- [20] A. M. Khan, Y.-K. Lee, S. Y. Lee, and T.-S. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 14, no. 5, pp. 1166–1172, 2010.
- [21] J.-Y. Yang, Y.-P. Chen, G.-Y. Lee, S.-N. Liou, and J.-S. Wang, "Activity recognition using one triaxial accelerometer: A neuro-fuzzy classifier with feature reduction," in *Entertainment Computing - ICEC 2007. 6th International Conference, Shanghai, China, September 15-17, 2007. Proceedings*, ser. Lecture Notes in Computer Science, L. Ma, M. Rauterberg, and R. Nakatsu, Eds., vol. 4740. Springer Verlag, 2007, pp. 395–400.
- [22] S.-W. Lee and K. Mase, "Activity and location recognition using wearable sensors," *IEEE Pervasive Computing*, vol. 1, no. 3, pp. 24–32, 2002, IEEE Computer Society Press, Los Alamitos, CA, USA, ISSN: 1536-1268.
- [23] C. Randell and H. Muller, "Context awareness by analyzing accelerometer data," in *ISWC '00: Proceedings of the Fourth IEEE International Symposium on Wearable Computers*. Washington, DC, USA: IEEE Computer Society Press, 2000, pp. 175–176.
- [24] N. Bhargava Bharatula and G. Tröster, "On-body context recognition with miniaturized autonomous sensor button," *Technisches Messen*, vol. 74, pp. 621–628, Dec. 2007, oldenbourg Wissenschaftsverlag GmbH, ISSN: 0171-8096.
- [25] G. S. Chambers, S. Venkatesh, G. A. W. West, and H. H. Bui, "Hierarchical recognition of intentional human gestures for sports video annotation," in *Proceedings of the 16th International Conference on Pattern Recognition*, vol. 2. IEEE Computer Society Press, 2002, pp. 1082–1085.
- [26] K. V. Laerhoven and O. Cakmakci, "What shall we teach our pants?" in *ISWC '00: Proceedings of the 4th IEEE International Symposium on Wearable Computers*. Washington, DC, USA: IEEE Computer Society Press, 2000, pp. 77–83.
- [27] J. Mantyjarvi, J. Himberg, and T. Seppanen, "Recognizing human motion with multiple acceleration sensors," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 2001*, vol. 2, 2001, pp. 747–752.
- [28] R. Pérez, Ú. Costa, M. Torrent, J. Solana, E. Opisso, C. Cáceres, J. M. Tormos, J. Medina, and E. J. Gómez, "Upper limb portable motion analysis system based on inertial technology for neurorehabilitation purposes," *Sensors*, vol. 10, no. 12, pp. 10733–10751, 2010.
- [29] M. Zhang and A. A. Sawchuk, "A feature selection-based framework for human activity recognition using wearable multimodal sensors," in *Proceedings of the 6th International Conference on Body Area Networks*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2011, pp. 92–98.
- [30] T. Plötz, N. Y. Hammerla, and P. Olivier, "Feature learning for activity recognition in ubiquitous computing," in *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Two*. AAAI Press, 2011, pp. 1729–1734.
- [31] XSens Technologies B.V., *MTw User Manual: MTw Hardware, MT Manager, Awinda Protocol*, 2013. [Online]. Available: <http://www.xsens.com/>
- [32] D. A. Winter, *Biomechanics and Motor Control of Human Movement*. Wiley, August 2004, ISBN: 978-0471509080.
- [33] K. P. Baclawski, "Bayesian network development," in *International Workshop on Software Methodologies, Tools and Techniques, pages 18-48. Keynote address.*, Sept 2004.
- [34] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco: Morgan Kaufmann, 1988.
- [35] F. V. Jensen and T. D. Nielsen, *Bayesian Networks and Decision Graphs*, 2nd ed. University of Aalborg, Department of Computer Sciences: Springer, 2007.
- [36] G. F. Cooper and E. Herskovits, "A bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 09, no. 4, pp. 309–347, Oct. 1992.
- [37] I. D. Coope and C. J. Price, "On the convergence of grid-based methods for unconstrained optimization," *SIAM Journal on Optimization*, vol. 11, no. 4, pp. 859–869, 2000, society for Industrial and Applied Mathematics, Philadelphia, PA, USA, ISSN: 1052-6234.