Recognition of Professional Activities With Displaceable Sensors

Dina Bousdar Ahmed*,Korbinian Frank†,Oliver Heirich*

*Institute of Communication and Navigation
German Aerospace Center - Munich (Germany)
{firstname.lastname}@dlr.de

†European Patent Office - Munich (Germany)

***Abstract*—The applications of activity recognition have increased in the last years. Among its uses are promoting healthy life styles or monitoring professional users (like first responders) during life-threatening operations; which caused the need for new research lines. However systems using common devices (like smart phones) equipped with the necessary sensors (e.g. inertial sensors) are not fully developed. Our work develops and activity recognition system, with a focus on professional users, that uses a displaceable sensor. The activity set is comprised of the most common activities (static, walking, running, etc.) but also crawling and 3D motions like stairs walking. A detector of upright dynamic activities has been developed and implemented to cope with the changing sensor position. The system is evaluated using 10-fold cross validation with the inference network, and recall with the detector. The final system proves to perform well with respect to the initial requirements.**

***Keywords*—activity recognition; motions; crawling; 3D activities; Bayesian; Grid-based filter; upright detector**

I. Introduction

Activity monitoring is a latent topic. Tracking a user's activity is extremely useful in several situations, i.e. in the healthcare context, either for inpatients or rehab patients at home. First responders often face risky situations in rescue missions, and monitoring their state is key to keep them safe. Along the last years many approaches have arisen to solve the problem. Yet a single solution cannot be provided since there are several degrees of freedom: the type, number and position of the sensing devices, the activity set, etc. Also targeting the mass market or professional applications entail different requirements. The possibility to employ daily-use devices in the monitoring is attractive because they are equipped with the appropriate sensors. We will be considering inertial measurement units (IMU) and a barometer for sensing. They are low cost, light weight, robust and getting more and more accurate every day. They can also be found in a wide variety of gadgets including smart phones, smart watches, tablets, etc.

An extensive analysis of the state-of-the-art was carried out. Although there were not two equal systems, some similarities could be identified. Smart phone-based systems, [1], frequently identify the same activities. They are equipped with additional tools (like GPS), which enables extending the activity set, e.g. to motorized vehicles [2]. We paid special attention to systems relying on wireless moveable sensors. Some interesting works were found, all with different approaches to estimate the sensor position. One option is using additional sensors, like light or proximity sensors, [3]. Another one is limiting the activity set, as in [4]. Other works, like [5], develop a complex algorithm to keep track of the sensor orientation. To the best of our knowledge, there is still no activity recognition system using displaceable sensors (with a limitation to inertial sensors, magnetometers and barometer) that targets a wide activity set.

The purpose of this work is to develop and assess the performance of an activity recognition system for professional users using a displaceable sensing platform. Professional users, i.e. first responders, cannot afford in some scenarios activities like walking or running. Thus they have to switch, for example, to activities such as high crawling. To consider these use cases, the activity set comprises static activities like *standing*, *sitting*, *lying*, *elevator up* and *elevator down*, and dynamic activities like *walking*, *walking up stairs*, *walking down stairs*, *running*, *jumping*, *falling*, *high crawling* and *low crawling*.

II. Approach

We use a sensing platform placed in one of several body positions. Table I shows the positions considered for each activity in the set. Sensor positions were chosen to emulate the use of a smart phone as sensing device. Therefore positions such as *texting* or *phoning* have not being considered with activities like *jumping*.

TABLE I. Reasonable pairs activities-sensor position

	Belt	*Pocket*	*Hand*	*Texting*	*Phoning*
Standing	×	×	×	×	×
Sitting	×	×	×	×	×
Lying	×	×	×	×	×
Elevator up	×	×	×	×	×
Elevator down	×	×	×	×	×
Walking	×	×	×	×	×
Walking up stairs	×	×	×	×	×
Walking down stairs	×	×	×	×	×
Running	×	×	×	×	×
Jumping	×	×	×		
Falling	×	×	×	×	×
High crawling	×	×			
Low crawling	×	×			

Since the sensor position can change while the activity is monitored, the measurements refer to both the targeted activity set and changes in sensor position. This is not critical for dynamic activities because they entail higher accelerations and turn rates than switching the sensor position. However, static activities are characterized by assessing the user's torso

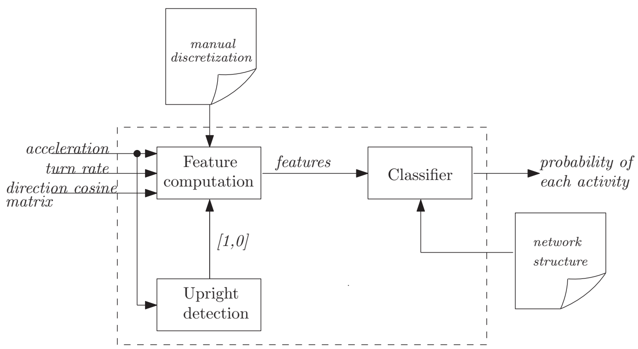

Fig. 1. Block diagram of the activity recognition system.

position with respect to gravity. Changes in sensor position also change the reference of the user's body, so the latter has to be updated. Figure 1 depicts the approach. Sensor data is processed to compute a set of features. These are used by the classifier to compute the probability of each activity. In parallel, acceleration is analysed to detect upright dynamic activities. If detected, the body reference is updated. Upright dynamic activities include *walking*, *walking up stairs*, *walking down stairs*, *running* and *jumping*.

It is important to highlight that this system provides the most likely activity regardless of the sensor position. The latter is never estimated. The main advantage of this approach is that it limits the effect of the changing position in the accuracy of recognition.

III. Data set

Record a suitable data set is the first task to develop this system. The data set must be representative of the population so it should balance male and female volunteers, age distribution and as many ways to perform the same activity as possible. Data from 19 different volunteers has been recorded under semi-naturalistic conditions. A wireless motion tracker (MTw) form Xsens is used [6], see Figure 2. The volunteers have been asked to perform a sequence of activities with the sensing platform placed in different positions. They were encouraged to perform the activities freely and in their own characteristic way. The data set has been labelled manually by a supervisor later on. Table II summarizes the characteristics of the data set. This data set is complemented with the one recorded in [7], which uses the same sensing device but a single sensor position (belt).

Fig. 2. MTw motion tracker (left) Awinda Dongle (right). The users carried the MTw and the Dongle was connected to a laptop to receive the data from the MTw.

TABLE II. Data set description

Number of volunteers	19
Time period for experiments	8 weeks
Age range	21-32
Total amount of data	8h 16' 42"

TABLE III. Set of features. Column *No.* represents the number of the feature. *Frame* gives the frame of representaion and *Definition* defines the feature. *Window* is given in number of samples

No.	Frame	Definition	Window
1	SF	$\lvert\bar{a}\rvert$	128
2		$E(\lvert\bar{a}\rvert)$	256
3		$\sigma(\lvert\bar{a}\rvert)$	256
4		$E(\lvert\bar{a}\rvert_{\text{BPF 1.6-4.5 Hz}})$	256
5		$\lvert\bar{a}\rvert$	32
6		$E_{SD2}(\lvert\bar{a}\rvert)$	512
7	GF	$\lvert\bar{a_h}\rvert$	128
8		$\sigma(\lvert\bar{a_h}\rvert)$	256
9		$\lvert\bar{a_h}\rvert$	32
10		$\sigma(MTZ(a_v))$	256
11		$RMS(MTZ(a_v))$	256
12		$E_{SD}(MTZ(\omega_{\text{h LPF}}))$	256
13	BF	$\bar{a}_h$	128
14		$\bar{a}_h$	256
15		$RMS(\bar{a}_h)$	256
16		$RMS(\bar{a}_h)$	384
17		$\bar{a}_v$	128
18		$RMS(\bar{a}_v)$	128
19	-	$SUM(Pol(RMS(p)))$	256
20		$IUD(p)$	128

IV. Features

Every activity has some physical properties that can be utilised to recognize it. The properties are used in the form of values, so called *features*. Using the latter instead of the raw data eases the task of the classifier and enhances the recognition of every activity. A feature can be defined as the result of a chain of mathematical operations; a description of features and the feature selection process can be found in [7]. It is relevant for our work to identify the main coordinate systems, also known as frames, where a feature can be represented. Initially, 3D-vectors measured by the sensor are given in the Sensor Frame (SF). Since this varies with the sensor orientation, a second fixed frame is defined: the Global Frame (GF). Yet the Body Frame (BF), defined according to the user's body, is the one that provides most information about the activity. Rotation from one frame to another is done using the direction cosine matrix (DCM), although other methods exist [8]. The inconvenience of the BF is the need for calibration, which is subjected to a certain sensor position. The selection of features is done assuming the correct calibration. How to achieve the latter will be addressed in a further section.

The classifier is discrete, so there is a need to turn the continuous value-range of a feature into a discrete one. Clustering techniques could be used, but we decided for a manual discretization using histograms and feature plots which gives more flexibility and a close to optimal separation of activity classes [7]. The following section describes the set selected.

A. Feature set

The feature selection process culminated in a set of 20 features, see Table III. Features are distributed along the three frames, and in the following we describe them according to the frame of representation. The description of the operators can be found in [7]. The activities in the figures are colour-coded according to the legend in Figure 3.

Fig. 3. Color codes for activities

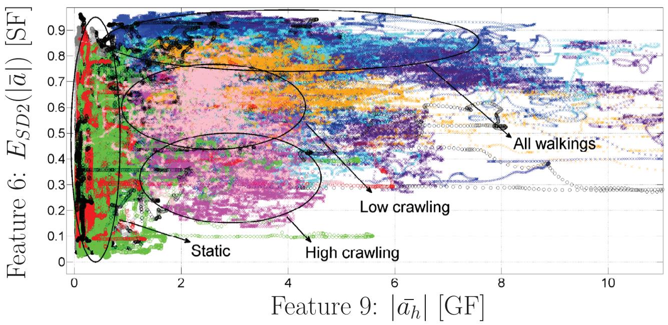

Fig. 4. Feature 9 versus feature 6

1) Features in the SF: since this frame varies with the sensor orientation, it is pointless to separate horizontal and vertical components. Instead, features in this frame are computed using the norm of 3D-vectors. This lets us identify three groups of activities: static (*standing*, *elevator up*, *elevator down*, *sitting* and *lying*), weakly dynamic (all *walkings*, *high crawling*, *low crawling* and *falling*) and highly dynamic (*running* and *jumping*). A relevant contribution of this work is the distinction of dynamically weak activities in the SF. This is done using feature 6, $E_{SD2}(|\bar{a}|)$. E_{SD2} is a feature that operates in the frequency domain. It computes two energies, each in a different frequency band, and normalizes one to the other. The energy of acceleration in 1 Hz-3.4 Hz is normalized to the energy between 0.5 Hz-10 Hz in a 5.12 s window. Since the *walkings* have more energy in the first band than any of the *crawlings*, activities are separated as in the y-axis of Figure 4.

2) Features in the GF: since this frame is fixed, a decomposition of 3D-vectors in horizontal and vertical components is possible. The DCM is provided by the MTw. Feature 11, $RMS(MTZ(a_v))$, is the root mean square of the vertical acceleration without gravity force. It focuses on the high vertical acceleration experienced by *jumping* while *running* reaches clearly lower values. In the y-axis of Figure 5 it can be seen how all dynamically weak activities occupy the same range, which is higher than that for static ones.

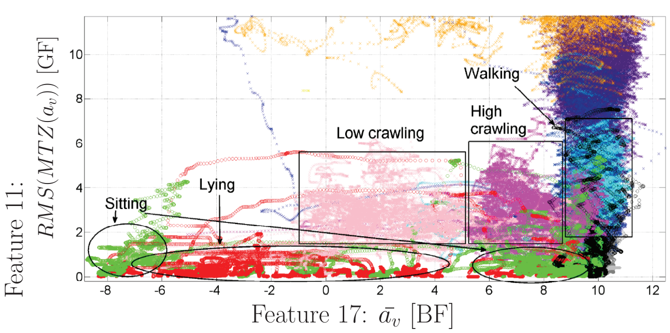

Fig. 5. Feature 17 versus feature 11

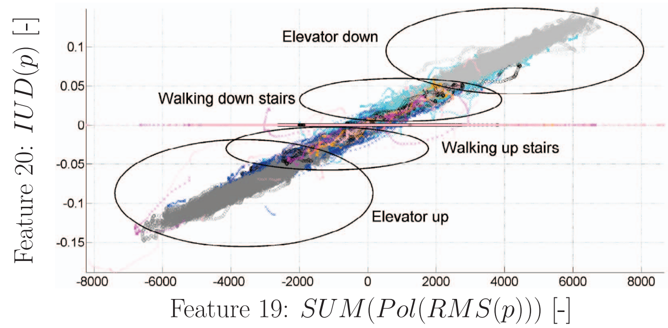

Fig. 6. Feature 19 versus feature 20

3) Features in the BF: are used mostly for static activities. These entail no relevant acceleration or turn rate, thus their recognition relies on assessing the BF with respect to the GF. Also the recognition of weakly dynamic activities can benefit from this analysis. Feature 17, $\bar{a}_v$, is the mean value of the vertical acceleration. Vertical torso activities are gathered around gravity force whereas horizontal torso activities are gathered around zero. Please note that *sitting* occupies two different ranges. The first between *standing* and *lying*, which is the expected one. The second nearby *lying*, with lower acceleration values. This happens when the sensor is placed in the pocket. If the latter is big, the sensor is on the thigh and during *sitting* it is completely horizontal. Thus confusing *sitting* with *lying*. The x-axis of Figure 5 also shows the separation between *low crawling*, *high crawling* and the *walkings*. All *walking* types are overlapping and they will be separated by further features.

4) Features without frame: are comprised of pressure-based features. They aim to identify motions in the vertical direction (*stairs walking* and *elevator up/down*) by looking for increasing or decreasing trends in pressure. In Figure 6 all activities with vertical motion are highlighted. *Walking up/down stairs* have smaller changes in pressure than standing in a moving elevator. The reason is that the elevator usually moves faster than a person walking the stairs. Hence, changes in pressure are higher when standing in the elevator. The pressure measured also depends on the atmospheric conditions, which can cause changes in pressure for activities with no height displacement, e.g. *low crawling*. This is why in Figure 6 there are different colours than those of *stairs walking* and *elevator up/down* out of the $(0, 0)$ zone.

V. Upright-Dynamic Detection

This section presents the upright detector that keeps track of the BF. A switch in the sensor position causes a loss of the reference of this frame. An update is needed to get the reference back, see section II. The update of the BF has to be done during upright activities, e.g. *standing*, *walking*, *running*, etc. In the case of *standing*, it can be prone to confusion with *sitting* and/or *lying* if the current BF reference is incorrect. Thus we decided to update the BF only for upright dynamic activities: *walking*, *walking up stairs*, *walking down stairs*, *running*, *jumping*.

The detection relies on analysing the frequency distribution of the acceleration norm. Figure 7 depicts the algorithm to detect upright dynamic activities. The norm of acceleration in the SF is used to make the detection independent of the sensor orientation. The energy in B_w (1 Hz-3.4 Hz) is compared to the energy in B_t (0.5 Hz-10 Hz), E_n in Figure 7. If this

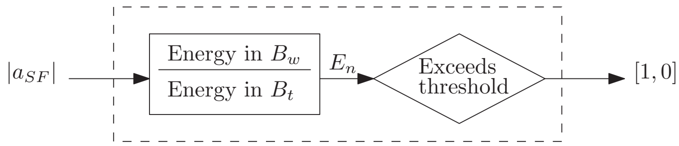

Fig. 7. Detection algorithm for upright dynamic activities. $|a_{SF}|$ is the norm of the acceleration vector in the SF. B_t and B_t are frequency bands. E_n represents the energy in B_w normalized to the energy in B_t. The energy is computed using Fast Fourier Transform coefficients. The output is 1 when the activity is an upright dynamic one or 0 otherwise.

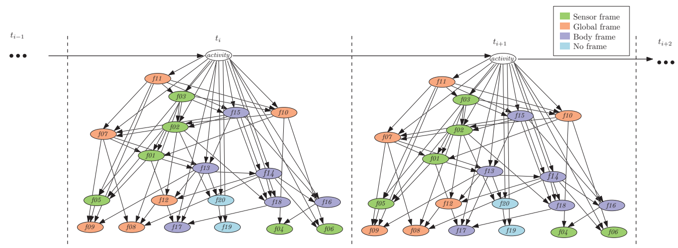

Fig. 8. Dynamic Bayesian network. Each structure at t_i is a Bayesian network. One ellipsoid represents a single random variable. The name of the latter refers to the feature number or the target random variable (*activity*). Each frame is represented by a colour. The dynamic nature is given by the edge that relates the activity node from one time to the next one.

value exceeds a threshold, a 1 is output meaning that the activity within the window is an upright dynamic one. This flag is used to update the BF, see the feedback from *upright detection* to *feature computation* in Figure 1. Otherwise a 0 is output and no action is taken. The selection of the parameters B_w, B_t, time window and threshold is done by analysing the frequency distribution of $|a_{SF}|$ for each activity. They are selected considering that static activities, *high crawling* and *low crawling* must not trigger an update.

VI. Inference Network

The feature values feed an inference network. The output is the most likely activity given the set of features. Among all possible classification techniques, we have decided for Bayesian methods. These can deal with missing or uncertain data. The output is the likelihood of each one of the possible states, not the most likely state. In this system a state is an activity. We decided for dynamic Bayesian networks (DBN). On the one hand, because a Bayesian approach proves to model better the relationship between the random variables (activity and features) than a naïve approach. On the other hand, because human motions are related over time and the use of a dynamic network can model this. The process of creating an inference network is known as *network learning* [9]. The learning machine used in this work is the same as the one used in [10].

The network was learnt after 15 days, with a limit of 4 parents per node. The structure is depicted in figure 8 at times t_i and t_{i+1}. The dynamic Bayesian network depicted in figure 8 is the first-order HMM used in this system. The *activity* node at time t_{i+1} is affected by the *activity* node at time t_i.

VII. Results

The performance of the classifier is assessed using 10-fold cross validation, [11]. The results are presented as the percentage of time an activity was classified as any of the 13 possible activities. Precision is the percentage of correct

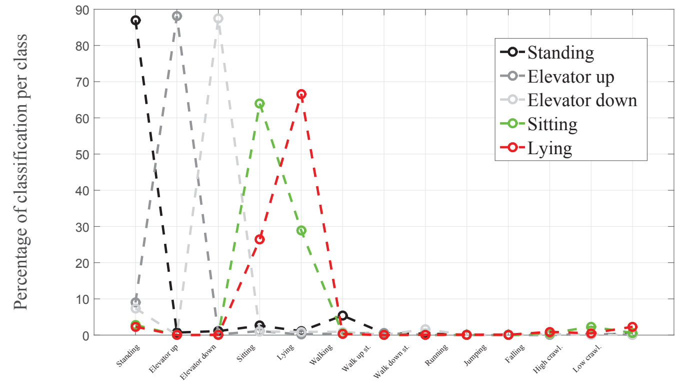

Fig. 9. Classification results for static activities

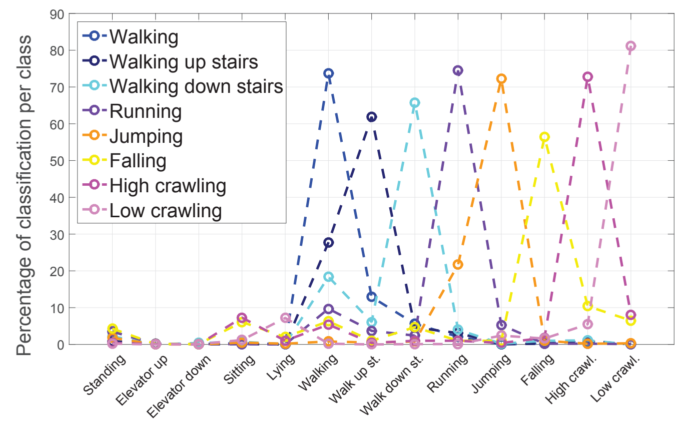

Fig. 10. Classification results for dynamic activities

classifications, and it is defined as follows:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

TP stands for true positives (correctly identified instances, e.g. when *running* is classified as *running*) and FP stands for false positives (retrieved instances that are false, e.g. if *running* is classified as *walking*, the FP of *walking* increase in one unit).

The data set used for validation was the one collected in Section III. It was split in 10 folds making sure that in all iterations the test set was comprised of all possible pairs activity-sensor position.

The upright-dynamic detector was evaluated using recall, see Equation (2). FN stands for false negatives (non-retrieved instances that are true, e.g. when if *running* is classified as *walking*, the FN of *running* increases in one unit).

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

The following sections present the results and some discussion about them.

A. Inference network

The results are summarized in Figures 9 and 10. Every activity was correctly classified more than 50% of the times. This is indicated by the precision, which corresponds to the maximum of each curve.

TABLE IV. RESULTS OF UPRIGHT-DYNAMIC DETECTION (%). *Activ.* IS SHORT FOR *activities*, *High c.* AND *Low c.* ARE SHORT FOR *high crawling* AND *low crawling* RESPECTIVELY

Activ.	*Static*	*Walk*	*Run*	*Jump*	*Fall*	*High c.*	*Low c.*
Recall	99.45	54.68	80.8	91.1	100	97.9	93.7

Regarding the group of static activities, *elevator up* and *elevator down* get the best precision. This is because of pressure-based features. If they are incorrectly classified, *standing* is the most likely activity they are confused with. This happens during the initial and last seconds in the elevator. At those times pressure does not vary significantly so they are misclassified as *standing*. *Sitting* and *lying* get the lowest values. They are confused with each other when the sensor is placed in the pocket, see Section IV. With respect to dynamic activities, we can see that *running* reports the best values. It is mostly confused with *walking* if the user is not running fast enough or with *jumping*, since this is also a phase of the *running* motion. See also how *jumping* is confused with *running* in 21.7% because they have similar acceleration values, specially in the SF. *High crawling* and *low crawling* precision values are over 70%; they are well characterized by their respective body positions. In the case of *stairs walking*, they hardly exceed 60% and are confused with *walking* in 27.7% of the times for *walking up stairs* and 18.4% of the times for *walking down stairs*. The reason is that if pressure does not sufficiently change in the window where pressure-based features are computed, their recognition is not triggered. Finally, *falling* is the activity with the worst results among the set. As explained in Section IV, short-activities are very hard to identify. Although features 5 and 9 are introduced, yet there are 18 features that look in longer windows. Thus resulting in a precision of only 56.4%.

B. Upright-dynamic detection

Table IV gives the recall for each activity. A TP in the case of upright-dynamic activities is 1 and a FN is a 0. For static activities, *high crawling* and *low crawling* a TP is a 0 and a FN a 1. *Static* refers to all static activities and *walking* refers to *walking*, *walking up stairs* and *walking down stairs*.

The results show good rejection of non-upright activities (static and crawling styles) with a minimum of 93.7% for *low crawling*. Regarding upright-dynamic activities, *running* and *jumping* reach recall values over 80%, which is reasonably high compared to the value reached by *walking*. Such low value, 54.68%, is obtained because the algorithm evaluates acceleration in both frequency distribution (via B_w) and energy level (via the threshold). In the case of *walking*, the frequency band is the right one. However if the user is not walking at a sufficiently high pace, E_n does not reach values that can exceed the threshold. Thus reporting a 0 instead of a 1 at the output of the detector.

VIII. CONCLUSION

Our work presents an activity recognition system with automatic update of the BF for upright-dynamic activities. Our system specially focuses on professional users. We use a displaceable sensor to emulate the use of common devices like smart phones. The system was evaluated using 10-fold cross validation for the inference network and recall for the detector.

The results show good recognition for dynamic activities regardless of the sensor position, because the motion characteristics are noticeable regardless of the sensor position. Static activities, however, show excessive reliance on the BF. Although the BF rotation works, sensor positions not attached to the body (*hand* or *texting*) cannot measure tilts in the torso. This makes the recognition of static activities even more complex. In general, activity recognition with displaceable sensors is possible under certain conditions. First, having the right rotation from the initial frame (SF) and the other two (GF and BF). Second, a valid rotation is not always possible in the case of the BF, specially for positions like *hand*, because a swinging hand moved independently of the body. In those cases some confusion must be admitted between static activities. Detection of upright-dynamic activities show promising results to compensate for the changing sensor position. Nevertheless the algorithm presented can be optimized, and opens a new line of research to enhance activity recognition with common devices. This system does not address the change in sensor position during static activities. We keep this as a future line of research along with the implementation and assessment of the system in a smart phone.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Stephan Sand, leader of the group of Vehicular Applications at the Institute of Communications and Navigation of the German Aerospace Centre, for his valuable suggestions and corrections.

REFERENCES

[1] C. A. Ronao and S.-B. Cho, "Human activity recognition using smartphone sensors with two-stage continuous hidden markov models," in *International Conference on Neural Computation (ICNC)*. IEEE, 2014, pp. 686–691.

[2] A. Anjum and M. U. Ilyas, "Activity recognition using smartphone sensors," in *Consumer Communications and Networking Conference CCNC*, no. 11-14. IEEE, January 2013, pp. 914–919.

[3] H. Martín *et al.*, "Activity logging using lightweight classification techniques in mobile devices," *Personal and Ubiquitous Computing*, no. 17, pp. 675–695, 2013.

[4] M. Susi, V. Renaudin, and G. Lachapelle, "Motion mode recognition and step detection algorithms for mobile phone users," *Sensors 2013*, vol. 13, no. 2, pp. 1539–1562, January 2013.

[5] V. Elvira *et al.*, "A novel feature extraction technique for human activity recognition," in *IEEE Workshop on Statistical Signal Processing*. IEEE, June 29-July 2 2014, pp. 177–180.

[6] X. Technologies, "Mtw user manual," pp. 14-18, 61.

[7] K. Frank *et al.*, "Bayesian recognition of safety relevant motion activities with inertial sensors and barometer," in *Position, Location and Navigation Symposium - PLANS 2014*. IEEE, May 5-8 2014, pp. 174–184.

[8] D. H. Titterton and J. L. Weston, *Strapdown Inertial Navigation Technology*. MIT Lincoln Laboratory copublished by the American Institute of Aeronautics and Astronautics, Inc, 2004.

[9] K. Backlawski, "Bayesian network development," in *New trends in software methodologies, tools and techniques*. IOS Press, 2004, pp. 18–48.

[10] K. Frank *et al.*, "Comparison of exact static and dynamic bayesian context inference methods for activity recognition," in *PERCOM Workshops*. IEEE, March 29 - April 2 2010, pp. 189–195.

[11] C. Shalizi, "Evaluating statistical models," 18 January 2011, http://www.stat.cmu.edu/~cshalizi/402/lectures/03-evaluation/lecture-03.pdf.