Automatic Recognition of Daily Living Activities Based on a Hierarchical Classifier

Oresti Banos, Miguel Damas, Hector Pomares, and Ignacio Rojas

Department of Computer Architecture and Computer Technology, University of Granada, C/Periodista Daniel Saucedo Aranda s/n E-18071, Granada, Spain
{oresti,mdamas,hpomares,irojas}@atc.ugr.es

Abstract. Physical activity recognition has become an increasing research area specially on health-related fields. The amount of different postures, movements and exercises in addition to the difficulty of the individuals particular execution style determine that extremely robust efficient knowledge inference systems are extremely necessary, being classification process one of the most crucial steps. Considering the power of binary classification in contrast to direct multiclass approaches, and the capabilities offered by multi-sense environments, we define a novel classification schema based on hierarchical structures composed by weighted decision makers. Remarkable accuracy results are obtained for a particular activity recognition problem in contrast to a traditional multiclass majority voting algorithm.

Keywords: hierarchical classification, weighted classification, binary classifiers, activity recognition.

1 Introduction

Over the past decade, there has been considerable research effort directed toward monitoring and classification of physical activity patterns from body-fixed sensor information. The application on several fields such as manufacturer industry [6], sports [2] or videogames interactive entertainment [11] is clearly recognized, however a special interest is recently being focused on healthcare. Chronic disease management [15], rehabilitation systems [10] or disease prevention [14] are several topics where activity recognition potential is being revealed.

One of the most important stages on activity recognition systems is machine learning. Several paradigms such as artificial neural networks [8], support vector machines [10], Bayesian classifiers [4] or hidden Markov models [12] have been widely used, but they are less accurate as the number of classes (activities) grows [9].

Some of these schemas are originally defined through binary classification, recognized as the most interesting approach [1], but in some cases traditional multiclass generalization is not efficiently practical. Besides, the use of several monitoring systems usually improves the system accuracy rates, but to the best of our knowledge, no general models are presented for the combined use of them. We

J. Cabestany, I. Rojas, and G. Joya (Eds.): IWANN 2011, Part II, LNCS 6692, pp. 185–193, 2011.

propose a wide-ranging multiclass schema by reducing the study to multiple binary or class specialized problems, employing a weighted structure to define the decision maker. This scheme is extended to each source to define a hierarchical knowledge inference system with a two-level weighting decision framework.

The rest of the paper is organized as follows. In section 2 a brief summary of the activity recognition process is presented. Section 3 describes the hierarchical weighted classification methodology proposed, showing the fundamentals of this method and the algorithm's main steps. Finally the performance of the method is evaluated for a specific example in section 4.

2 Activity Recognition Method

The experimental setup starts from a signal set [4] corresponding to acceleration values measured by a group of sensors located in several strategic body locations (hip, wrist, arm, ankle, thigh), for four daily activities (see Fig. 1). The methodology presented from this point forward can be easily generalized to other studies related to activity recognition from a set of features.

Monitored data have some artifacts and noise associated to the acquisition data process. Considering that a 20 Hz sampling is sufficient to assess habitual daily physical activity [7], a low pass elliptic filter with 20 Hz cutoff frequency, followed by a 0.5 Hz cutoff frequency high pass elliptic filter (both 0.5dB passband ripple and 20dB stopband attenuation) are applied to respectively remove the high frequency noise and the original signal offset.

Afterward a parameter set made up of 861 features is obtained. This corresponds to a combination of statistical functions such as mode, median, variance, etc., and magnitudes obtained from a domain transformation of the original data such as energy

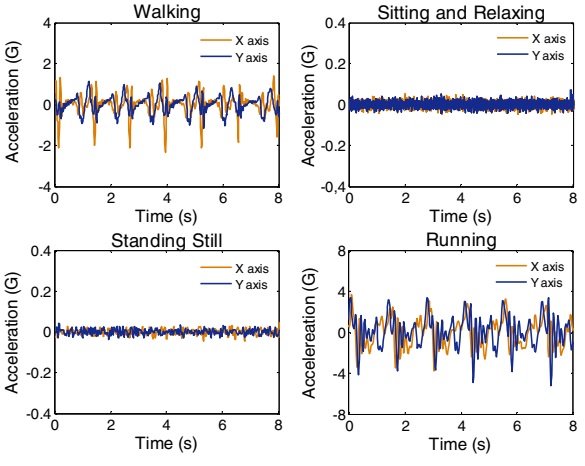

Fig. 1. Signals corresponding to four usual daily physical activities (ankle accelerometer). These activities have been considered for the daily living commonness, as well as because of the manner execution similarity between couples walking/running and sitting/standing.

Table 1. Feature set generation functions. Every statistical function is computed for each magnitude generating the set of 861 features.

Magnitudes	Statistical functions
Amplitude	4^{th} and 5^{th} central statistical moments
Autocorrelation function	Energy
Cepstrum	Arithmetic/Harmonic/Geometric/Trimmed mean
Correlation lags	Entropy
Cross correlation function	Fisher asymmetry coefficient
Energy spectral density	Maximum / Position of
Spectral coherence	Median
Spectrum amplitude/phase	Minimum / Position of
Histogram	Mode
Historical data lags	Kurtosis
Minimum phase reconstruction	Data range
Daubechies wavelet decomposition	Total harmonic deviation
	Variance
	Zero crossing counts

spectral density, spectral coherence or wavelet ("a1 to a5" and "d1 to d5" Daubechies levels of decomposition) among others for both signal axes. "Fisher asymmetry coefficient of the X axis signal histogram", "Y axis signal wavelet coefficients a2 zero crossing counts" or "X axis-Y axis cross correlation harmonic mean" are possible examples of features obtained from the set defined (Table 1).

Feature selection processes have the responsibility of deciding which features or magnitudes are the most important ones to infer the kind of activity the person is carrying out. Taking into account the binary class classifier approach (described in the next section), several class specialized feature selection schemas based on an 'one-against-all' strategy have been applied to the data. The methodologies implemented are based on Bhattacharyya distance, Entropy [13], and a technique (*Quality Group Ranking* or *QGR*) recently defined in a previous work [3].

3 Hierarchical Weighted Classifier (HWC)

Considering binary classification in general more accurate than direct multiclass approach, is extremely important to establish an appropriate multiclass extension schema which permits to preserve and optimize binary entities capabilities, even more when fusion of several sensors or sources is considered. A general methodology based on the combination of binary or class classifier decisions in a hierarchical structure with an special application for multisource problems is presented in this section.

The framework of the HWC is composed by three classification levels or stages related to the decision structure defined (see Fig. 2). In general, for $m=1,...,M$ sources and $n=1,...,N$ classes, a set of M x N "class classifiers" (c_{mn}) are defined. They are binary classifiers specialized in the classification of the class n by using the data

acquired from the *m* source. Each one apply an 'one-versus-rest' strategy, so any classification paradigm can be easily applied. These define the first level or *class level classifier*. The second stage, *source level classifier*, is defined by *M* "source classifiers" (S_m). Source classifiers are not machine learning as class classifiers, but hierarchical decision models which define a classification entity. Source classifiers structures are composed by several class classifiers as is shown in Fig. 2, defining a decision system based on class classifiers weighted decisions. This approach is repeated for the next level, *method level classifier*, which ultimately defines a decision structure constituted by source classifiers weighted decisions.

In accordance to the structure described above, a process consisting of a few main steps is carried out to define the complete HWC. The process starts by evaluating the individual accuracy of each class classifier, defined through its corresponding feature vector (several vector lengths should be considered to find out the best results for every classifier). A 10-fold cross validation is suggested for accomplish this task and this is repeated 100 times to ensure the statistical robustness. The entirely process is repeated for each source. Considering average accuracy rates ($\overline{R_{mn}}$ for source *m* and class classifier *n*) as a measure of the pattern recognition capabilities of each classifier, an associated weight is obtained for each one:

$$\beta_{mn} = \frac{\overline{R_{mn}}}{\sum_{k=1}^{N} \overline{R_{mk}}} \tag{1}$$

These weights are a measure of the importance that every class classifier will have on the source classifier decision schema. An specific voting algorithm is considered in this point to define source classifiers decision. For a source *m*, given a sample x_{mk} to

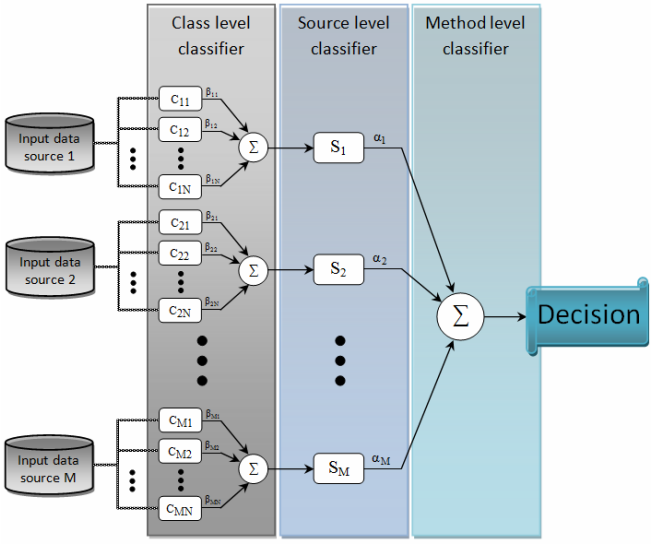

Fig. 2. HWC general structure for a problem with N classes and M sources

be classified and being q the class predicted by the classifier c_{mn}, if the sample is classified as belonging to the classifier class of specialization ($q=n$), the classifier will set its decision as '1' for the class n and '0' for the rest of classes. Opposite is made for ($q\neq n$). In summary, the decision from the classifier n for the class q is:

$$D_{nq}(x_{mk}) = \begin{cases} 1, & x_{mk} \ classified \ as \ q \\ 0, & x_{mk} \ not \ classified \ as \ q \end{cases} \forall q = n \qquad \begin{cases} 1, & x_{mk} \ not \ classified \ as \ q \\ 0, & x_{mk} \ classified \ as \ q \end{cases} \forall q \neq n \qquad \forall q = 1,..., N \quad (2)$$

Once decisions have been offered by each class classifier for every class q by applying (2), it is time to compute the weighted output for the m source classifier:

$$O_{mq}(x_{mk}) = \sum_{n=1}^{N} \beta_{mn} D_{nq}(x_{mk}) \qquad \forall q = 1,..., N \quad (3)$$

Finally, the class predicted for the m source classifier (q_m) is the class q for which source classifier output is maximized:

$$q_m = \arg \max_q \left(O_{mq}(x_{mk}) \right) \quad (4)$$

In this stage, the source level classifier is completely defined. Every source class classifier can be used separately, looking for the most interesting for the particular problem analyzed. If extremely accurate classifiers are found, maybe this is enough to be used as the final pattern recognition system solution. However, *fusion* or combination of sources information is in general a more robust and efficient solution. Consequently, the complete process described before is extended to a new hierarchy level, the *method level classifier*. First, source classifiers weights (α_m) are obtained by calculating the average accuracy rates for each source classifier ($\overline{R_m}$), so a cross validation process is again repeated but now focusing on the source classifiers predictions. The weight for the source m is:

$$\alpha_m = \frac{\overline{R_m}}{\sum_{k=1}^{M} \overline{R_k}} \quad (5)$$

The output is calculated taking into account the individual outputs obtained for each source classifier. For a sample x_k defined through the corresponding information obtained from each source ($x_{1k},...,x_{Mk}$):

$$O_q(x_k) = O_q(\{x_{1k},..., x_{Mk}\}) = \sum_{p=1}^{M} \alpha_p O_{pq}(x_{pk}) \qquad \forall q = 1,..., N \quad (6)$$

Similar to (4) the final class predicted q is:

$$q = \arg\max_{q} \left(O_q(x_k) \right) \qquad q \in [1,..., N] \tag{7}$$

In summary, the HWC is absolutely defined through the class classifiers (c_{mn}), class level weights (β_{mn}) and the source level weights (α_m) at this point.

4 Results

The aim of the methodology presented is to define robust and efficient pattern recognition systems based on binary class classifiers. For the activity recognition problem presented (N = 4, M = 5), two classification schemas based on majority voting (MV) and our approach (HWC) are respectively used. Naive Bayes machine learning paradigm is applied for the structure of class classifiers, using the two best features selected by each feature selector (for every source and class, see table 2), assuming the stochastical independence is satisfied. Fig. 3 shows the results.

Clearly, a comparison of MV and HWC allows us to realize the potential of HWC. All MV average accuracy rates for source classifiers are improved significantly by using HWC, up to nearly 35% in some cases as *ankle* source for QGR features selected. Furthermore, this is extensible to *fusion* of source classifiers or *method classification*, which are considerable better in all cases. In fact, no improvement appears for *fusion* approach when is used following a MV schema. Conversely, important benefits are obtained for HWC, with particular remarkable results for Bhattacharyya and Entropy, with accuracy rates above 90%. Because of QGR source classifiers are extremely accurate (>96%), results for *fusion* is in line to them (~99%).

As was mentioned in section 3, if source classifiers offer outstanding accuracy results, fusion approach may be omitted. This can be seen in Fig.3b, for QGR case. Source classifiers as based on the *thigh* or the *wrist* accelerometer define almost absolute recognition systems (~100%), so for this problem it would not be necessary to use other sensors, something especially important in wearable monitoring contexts.

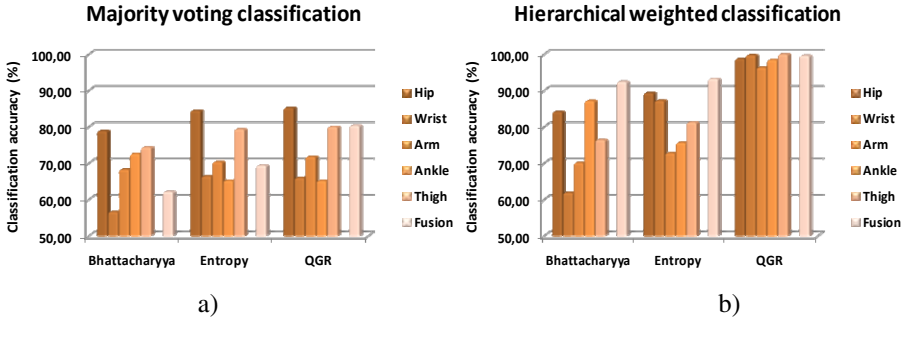

Fig. 3. Accuracy rates using a) MV and b) HWC. Results for each source classifier are identified with the corresponding sensor label. Fusion is referred to the combined use of the different source classifiers (identified as *method level classifier* in section 3).

Table 2. Two best features selected by each feature selection schema used. The features are used for the corresponding class classifier defined through the accelerometer and activity of specialization.

		Walking	*Sitting*	*Standing*	*Running*
BHATTACHARYYA	*Hip*	*Y-axis autocorr. function 5th central moment & Y-axis energy spectral density 5th central moment*	*X-axis minimum phase reconstruction position of the maximum & X-axis energy spectral density 5th central moment*	*X-axis minimum phase reconstruction position of the maximum & Y-axis minimum phase reconstruction position of the minimum*	*Y-axis autocorr. function 5th central moment & Y-axis energy spectral density 5th central moment*
	Wrist	*Y-axis autocorr. function 5th central moment & Y-axis autocorr. function 4th central moment*	*X-axis minimum phase reconstruction position of the maximum & X-axis energy spectral density 5th central moment*	*Y-axis autocorr. function entropy & Y-axis autocorr. function position of the minimum*	*Y-axis autocorr. function 5th central moment & Y-axis energy spectral density 5th central moment*
	Arm	*X-axis autocorr. function 5th central moment & Y-axis autocorr. function 5th central moment*	*Y-axis energy spectral density 5th central moment & Y-axis autocorr. function 5th central moment*	*X-axis autocorr. function entropy & X-axis autocorr. function position of the minimum*	*X-axis minimum phase reconstruction position of the maximum & X-axis autocorr. function 5th central moment*
	Ankle	*X-Y-axes cross correlation function 5th central moment & Y-axis autocorr. function 5th central moment*	*X-axis energy spectral density 5th central moment & X-Y-axes cross correlation function 5th central moment*	*X-axis autocorr. function entropy & X-axis autocorr. function position of the maximum*	*X-axis minimum phase reconstruction position of the maximum & X-Y-axes cross correlation function 5th central moment*
	Thigh	*X-axis minimum phase reconstruction position of the maximum & X-axis autocorr. function 5th central moment*	*Y-axis energy spectral density 5th central moment & Y-axis autocorr. function 5th central moment*	*X-axis autocorr. function entropy & X-axis autocorr. function position of the minimum*	*X-axis autocorr. function 5th central moment & X-axis autocorr. function 4th central moment*
ENTROPY	*Hip*	*Y-axis amplitude trimmed mean & Y-axis amplitude median*	*X-axis amplitude zero crossing counts & X-axis amplitude standard deviation*	*X-axis histogram standard deviation & X-axis histogram maximum*	*X-axis amplitude zero crossing counts & X-axis amplitude standard deviation*
	Wrist	*X-axis amplitude trimmed mean & X-axis amplitude median*	*X-axis amplitude zero crossing counts & X-axis amplitude standard deviation*	*Y-axis spectrum amplitude Fisher asymmetry coef. & Y-axis spectrum amplitude kurtosis*	*X-axis amplitude standard deviation & X-axis amplitude maximum*
	Arm	*X-axis wav. a3 coef. zero crossing counts & Y-axis wav. d2 coef. zero crossing counts*	*X-axis amplitude standard deviation & X-axis amplitude energy*	*X-axis spectrum amplitude Fisher asymmetry coefficient & X-axis spectrum amplitude kurtosis*	*X-axis amplitude zero crossing counts & X-axis amplitude standard deviation*
	Ankle	*Y-axis wav. d1 coef. zero crossing counts & Y-axis wav. d3 coef. zero crossing counts*	*X-axis amplitude standard deviation & X-axis amplitude energy*	*X-axis amplitude kurtosis & X-axis histogram Fisher asymmetry coef.*	*X-axis amplitude zero crossing counts & X-axis amplitude standard deviation*
	Thigh	*X-axis wav. d1 coef. zero crossing counts & Y-axis autocorr. function zero crossing counts*	*X-axis amplitude zero crossing counts & X-axis amplitude standard deviation*	*X-axis minimum phase reconstruction position of the minimum & Y-axis amplitude minimum*	*X-axis amplitude standard deviation & X-axis amplitude energy*
QGR	*Hip*	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*
	Wrist	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*
	Arm	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*
	Ankle	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*
	Thigh	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*	*X-axis wav. a3 coef. geometric mean & Y-axis wav. a3 coef. geometric mean*

5 Conclusions

Several advantages are obtained to traditional multiclass schemas as majority voting. Primarily only features with high binary discriminant capacity are required because of class specialized classifiers define completely the knowledge base of the model. This reduces the complexity of feature selection processes. Besides, once source and class level weights are calculated for the corresponding problem analyzed, the classification system is simply defined through a few decision rules that are easily extended from source classifiers to the complete hierarchy.

A simple activity recognition system has been defined by using solely two features for each class classifier with accuracy rates close to 100% for some cases. Results are particularly interesting for the first level of the hierarchy defined (source classifier), having remarkable accuracy rates for some specific sensors as the wrist (specially for QGR system defined), so interesting for the unobtrusively and applicability on wearable monitoring activity recognition systems.

The good results obtained for the example above are promising for applying this technique to a problem with more classes. For future work we want to test our methodology in different related problems or others (*UCI repository* [5]) with a spread range of classes.

Acknowledgments. We want to express our gratitude to Prof. Stephen S. Intille, Technology Director of the House_n Consortium with the MIT Department of Architecture for the experimental data provided. This work was supported in part by the Spanish CICYT Project TIN2007-60587, Junta de Andalucia Projects P07-TIC-02768 and P07-TIC-02906, the CENIT project AmIVital, of the "Centro para el Desarrollo Tecnológico Industrial" (CDTI- Spain) and the FPU Spanish grant AP2009-2244.

References

1. Allwein, E.L., Schapire, R.E., Singer, Y.: Reducing multiclass to binary: a unifying approach for margin classifiers. J. Mach. Learn. Res. 1, 113–141 (2001)
2. Baca, A., Dabnichki, P., Heller, M., Kornfeind, P.: Ubiquitous computing in sports: A review and analysis. Journal of Sports Sciences 27, 1335–1346 (2009)
3. Banos, O., Pomares, H., Rojas, I.: Ambient living activity recognition based on feature-set ranking using intelligent systems. In: The 2010 International Joint Conference on Neural Networks (IJCNN), pp. 1–4 (2010)
4. Bao, L., Intille, S.: Activity Recognition from User-Annotated Acceleration Data. Pervasive Computing, 1–17 (2004)
5. Frank, A., Asuncion, A.: UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine, CA (2010), `http://archive.ics.uci.edu/ml`
6. Koskimaki, H., Huikari, V., Siirtola, P., Laurinen, P., Roning, J.: Activity recognition using a wrist-worn inertial measurement unit: A case study for industrial assembly lines. In: 17th Mediterranean Conference on Control and Automation MED 2009, pp. 401–405 (2009)

7. Mathie, M.J., Coster, A.C.F., Lovell, N.H., Celler, B.G.: Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. Physiol. Meas. 25, 1–20 (2004)
8. Parkka, J., Ermes, M., Korpipaa, P., Mantyjarvi, J., Peltola, J., Korhonen, I.: Activity classification using realistic data from wearable sensors. IEEE Transactions on Information Technology in Biomedicine 10, 119–128 (2006)
9. Preece, S.J., Goulermas, J.Y., Kenney, L.P.J., Howard, D., Meijer, K., Crompton, R.: Activity identification using body-mounted sensors—a review of classification techniques. Physiol. Meas. 30, 1–33 (2009)
10. Sazonov, E., Fulk, G., Sazonova, N., Schuckers, S.: Automatic Recognition of postures and activities in stroke patients. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBC 2009, pp. 2200–2203 (2009)
11. Schlömer, T., Poppinga, B., Henze, N., Boll, S.: Gesture recognition with a Wii controller. In: Proceedings of the 2nd international conference on Tangible and embedded interaction - TEI 2008 (2008)
12. Singla, G., Cook, D., Schmitter-Edgecombe, M.: Incorporating temporal reasoning into activity recognition for smart home residents. In: AAAI Workshop - Technical Report WS-08-11, pp. 53–61 (2008)
13. Theodoridis, S., Koutroumbas, K.: Pattern Recognition, 4th edn. Elsevier, Amsterdam (2009)
14. Warren, J.M., et al.: Assessment of physical activity – a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. European Journal of Cardiovascular Prevention & Rehabilitation 17, 127–139 (2010)
15. Zwartjes, D., Heida, T., van Vugt, J., Geelen, J., Veltink, P.: Ambulatory Monitoring of Activities and Motor Symptoms in Parkinson's Disease. IEEE Transactions on Biomedical Engineering 57, 2778–2786 (2010)