Towards Big Data for Activity Recognition: A Novel Database Fusion Strategy

Dominik Schuldhaus
Digital Sports Group
Pattern Recognition Lab
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Germany
dominik.schuldhaus@fau.de

Heike Leutheuser
Digital Sports Group
Pattern Recognition Lab
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Germany
heike.leutheuser@fau.de

Bjoern M. Eskofier
Digital Sports Group
Pattern Recognition Lab
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Germany
bjoern.eskofier@fau.de

ABSTRACT

Activity recognition is mandatory in order to provide feedback about the individual quality of life. Usually, activity recognition algorithms are evaluated on one specific database which is limited in the number of subjects, sensors and type of activities.

In this paper, a novel database fusion strategy was proposed which fused three different publicly available databases to one large database consisting of 42 subjects. The fusion of databases addresses the two attributes high volume and high variety of the term "big data". Furthermore, an algorithm was developed which can deal with multiple databases varying in the number of sensors and activities. Nine features were computed in sliding windows of inertial data of several sensor positions. Decision-level fusion was performed in order to combine the information of different sensor positions.

The proposed classification system achieved an overall mean classification rate of 85.8 % and allows an easy integration of new databases. Using big data is necessary to develop robust and stable activity recognition algorithms in the future.

Categories and Subject Descriptors

I.5.4 [**Pattern Recognition**]: Applications

General Terms

Algorithms

Keywords

Activity recognition, big data, database fusion, data mining, decision-level fusion, inertial sensors

BODYNETS 2014, September 29-October 01, London, Great Britain

DOI 10.4108/icst.bodynets.2014.256946

1. INTRODUCTION

According to the World Health Organization, physical inactivity has been identified as the 4th leading risk factor for global mortality [21]. Approximately 3.2 million people of the world population decease each year due to a physical inactive lifestyle [21]. Furthermore, physical inactive people suffer from breast and colon cancer, diabetes and ischaemic heart disease [21]. Compared to physically inactive people, physically active people have higher levels of health-related fitness and lower rates of various chronic diseases [12].

Methods for the assessment of the individual daily life physical activity support the monitoring of the health status and can be used for providing feedback about the individual quality of life. An overview of methods for the assessment of daily life activities is given for example in [20]. The most widely used tools to assess physical activity are self-report instruments including self or interviewer-administered questionnaires, recalls and activity diaries [16]. However, self-reports induce problems with reliability, validity and sensitivity [18]. In contrast to self-reports, small and light-weight wearable sensors like inertial measurement units provide a more reliable and objective assessment of physical activity. One common application of these wearable sensors is activity recognition. By recognizing daily life activities like walking, running, vacuuming or climbing stairs, feedback about the human's behavior is provided.

In order to build up robust and stable activity recognition systems, the algorithms should be trained on a large amount of data. The training on a large database offers flexibility to support new users without the need of collecting additional training data [9].

Mobile computing clouds enables researchers for computing on collections of large-scale sensor data, coming from embedded sensors in mobile devices, and support user feedback based on the analysis of big data [8]. The definition of "big data" focuses on the three Vs volume, variety, and velocity [15].

The attribute volume refers to the quantitative characteristics of big data. Regarding activity recognition, big data might be quantified by the number of available datasets.

The attribute variety addresses the diversity in terms of sources, data types, and entities. Regarding activity recognition, big data might include different sensor positions, sensor orientations, sensor types, types of activities, and execution of activities.

The attribute velocity refers to the frequency of data generation or the data delivery. Regarding activity recognition, big data might refer to the real-time data streams coming from any kind of device or sensor. In the following paragraphs, examples for big data are given.

In [5], an activity information gathering system was developed using mobile sensor devices with triaxial accelerometers. Inertial data of 170 people during five months were acquired. The study protocol included daily life activities like walking, taking a car, or changing clothes.

In [2], the challenge of different sensor orientations was addressed. A statistical analysis of acceleration features quantified relative effects of ideal, self-placement of a user, and mutual displacement deployments. The analysis was based on 33 fitness activities, recorded in a cardio-fitness room using nine inertial-magnetic measurement units from 17 participants.

In [14], a large scale multimodal data set of naturalistic human activities in a sensor rich environment was described. The sensor setup included 72 sensors of 10 modalities, e.g. microphone and inertial-magnetic measurement unit, integrated in the environment, in objects, and on the body.

In [7], the Human Activity Sensing Consortium was introduced. The aim was to collect a large scale human activity corpus of different research groups. By the end of 2010, more than 6700 accelerometer data of 540 subjects were available including the activities walking, jogging, jumping, and climbing stairs.

In [5, 2, 14], big data was generated using their own study reaching a high number of subjects [5], a high variety of sensor orientations [2], and a high variety of sensor types [14]. For future research, it is mandatory to further generate big data by combining different databases. The database fusion exploits the advantages of each study design. The Human Activity Sensing Consortium collects inertial sensor data of different research groups but the number of sensor positions is restricted to one [7]. The optimal sensor position depends on the type of activities to be classified [9]. Therefore, the fusion of different sensor positions should be considered for the recognition of a broad range of daily life activities.

Thus, the purpose of this paper was twofold. First, a novel database fusion strategy is introduced addressing the two attributes volume and variety of big data. The database fusion strategy increased the amount of data containing different sensor positions, sensor types, and types of activities. The required and necessary steps were described which were needed to combine multiple databases. The strategy was applied to three publicly available databases [13, 22, 10] containing inertial sensor data. To the best of the authors' knowledge, the idea of fusing sensor-based activity recognition databases has not yet been considered by research groups.

Second, the large database was used for the evaluation of an activity recognition algorithm based on decision-level fusion combining the information of different sensor positions. The proposed approach achieved an overall mean classification rate of 85.8 %.

In the future, database fusion strategies are mandatory, since the amount of data and the variety in the data can be increased. This idea of big data offers the possibility to further increase the performance of activity recognition algorithms.

2. MATERIALS AND METHODS

2.1 Datasets

In this paper, three publicly available databases were combined to one large database. In the following section, the corresponding databases are introduced.

2.1.1 PAMAP2 Physical Activity Monitoring Dataset

In the PAMAP2 dataset, three inertial-magnetic measurement units consisting of two triaxial accelerometers, a triaxial gyroscope and a triaxial magneto-resistive magnetic sensor were used [13]. The range of the accelerometer, gyroscope and magnetometer was $\pm 16g$ / $\pm 6g$, $\pm 1500°/s$ and $\pm 400\mu T$, respectively. The sampling rate was 100 Hz. The sensors were placed on the chest, wrist on dominant arm and ankle on dominant side. Furthermore, a heart rate monitor was used.

The dataset included nine subjects (1 female and 8 male, age 27.2 ± 3.3 years, BMI 25.1 ± 2.6 kgm^{-2}). One subject was left-handed, all the others were right-handed. Each subject had to perform 12 activities (lying, sitting, standing, walking, running, cycling, Nordic walking, watching TV, computer work, car driving, ascending stairs, descending stairs, vacuuming, ironing, folding laundry, house cleaning, playing soccer, rope jumping). The dataset can be downloaded from *http://www.pamap.org/demo.html*.

2.1.2 University of Southern California Human Activity Dataset (USC-HAD)

In the USC-HAD dataset, a single inertial measurement unit consisting of a triaxial accelerometer and a triaxial gyroscope was used [22]. The range of the accelerometer and gyroscope was $\pm 6g$ and $\pm 500°/s$, respectively. The sampling rate was 100 Hz. The sensor was placed on the right hip.

The dataset included 14 subjects (seven male and seven female, age 30.1 ± 7.2 years, height 170 ± 6.8 cm, weight 64.6 ± 12.1 kg). Each subject had to perform 12 activities (walking forward, walking left, walking right, walking upstairs, walking downstairs, running forward, jumping, sitting, standing, sleeping, elevator up, elevator down). Day-to-day activity variations were considered by performing five trials for each activity on different days at various indoor and outdoor locations. The dataset can be downloaded from *http://sipi.usc.edu/HAD/*.

2.1.3 Daily Life Activities (DaLiAc) Dataset

In the DaLiAc dataset, four inertial measurement units each consisting of a triaxial accelerometer and a triaxial gyroscope were used [10]. The sensors were placed on the right hip, chest, right wrist, and left ankle. The range of the accelerometer was $\pm 6g$ for all four sensor positions. The range of the gyroscope was $\pm 500°/s$ for the wrist, chest and hip position and $\pm 2000°/s$ for the ankle position. The sampling rate was 204.8 Hz.

The dataset included 19 subjects (8 female and 11 male, age 26 ± 8 years, height 177 ± 11 cm, weight 75.2 ± 14.2 kg). Each subject had to perform 13 activities (sitting, lying, standing, washing dishes, vacuuming, sweeping, walking, ascending stairs, descending stairs, running on treadmill, bicycling on ergometer (50 and 100 watt), rope jumping). The dataset can be downloaded from *http://www.activitynet.org*.

Table 1: List of selected activities and the corresponding abbreviations. The availability of the activities regarding the three databases is indicated by 'x'.

Activity	Abbreviation	USC-HAD	DaLiAc	PAMAP2
Static	ST	x	x	x
Walking	WK	x	x	x
Climbing stairs	CS	x	x	x
Running	RU	x	-	x
Jumping	JP	x	-	-
Vacuuming	VC	-	x	x
Bicycling	BC	-	x	x
Rope jumping	RJ	-	x	x

Table 2: Selected sensor positions and sensor types, the corresponding abbreviations and the corresponding number of available subjects.

Sensor position	Sensor type	Abbreviation	# Subjects
Wrist	Accelerometer	WR - ACC	28
Wrist	Gyroscope	WR - GYR	28
Chest	Accelerometer	CH - ACC	28
Chest	Gyroscope	CH - GYR	28
Hip	Accelerometer	HP - ACC	33
Hip	Gyroscope	HP - GYR	33
Ankle	Accelerometer	AK - ACC	28
Ankle	Gyroscope	AK - GYR	28

2.2 Database Fusion

For database fusion, six steps are required. They are described in the following section.

2.2.1 *Activity Selection*

The first step was to select the desired activities that should be classified. In this paper, eight activities were chosen that typically appear in daily life. The selected activities are described below.

- **Static**: sitting and standing were available in all three databases. Lying was available in PAMAP2 and DaLiAc and sleeping in USC-HAD. These activities were merged, since in many applications the posture of the body is not important.
- **Walking**: this was available in all three databases.
- **Climbing stairs**: ascending and descending stairs were available in all databases and were merged, since the direction can straightforwardly be assessed for example by integration of the up/down acceleration component.
- **Running**: running outside was only available in the USC-HAD and PAMAP2 database.
- **Jumping**: this was only available in the USC-HAD database.
- **Vacuuming**: this was only available in the DaLiAc and PAMAP2 database.
- **Bicycling**: this was only available in the DaLiAc and PAMAP2 database. In the DaLiAc database, two resistance levels on a stationary bike, namely 50 and 100 watt, were available. In the PAMAP2 database, cycling outside was available. In order to consider a high variety of different bicycling conditions, all mentioned activities were merged.
- **Rope jumping**: this was only available in the DaLiAc and PAMAP2 database.

Table 1 shows the eight selected activities, the corresponding abbreviations and the availability regarding the three databases.

2.2.2 *Sensor Position and Type Selection*

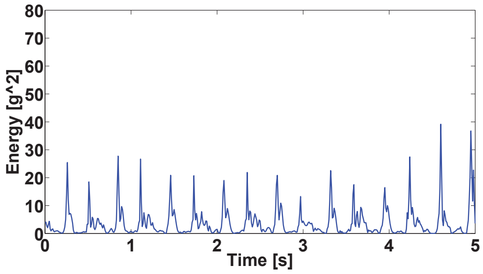
The second step was to select the desired sensor position that should be considered for activity recognition. In this paper, all available sensor positions were used, in order to acquire data of different body parts. Furthermore, the desired sensor types (accelerometer, gyroscope, magnetometer and heart rate monitor) had to be selected. In this paper, accelerometer and gyroscope were used, since they were also mostly used in literature [23, 10]. Table 2 shows the number of available subjects regarding each sensor position and type.

2.2.3 *Unit Adjustment*

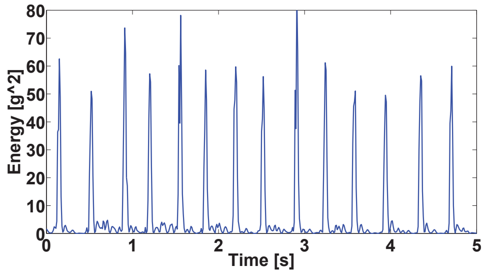
The third step was to adjust the unit of each sensor type. Since all databases used g as unit for the accelerometer, it remained unchanged. The PAMAP2 database used rad/s as unit for the gyroscope. USC-HAD and DaLiAc used $^{\circ}/s$ as unit for the gyroscope. Thus, the angular velocities in the PAMAP2 database were converted to $^{\circ}/s$.

2.2.4 *Normalization*

The fourth step was to define the common range of amplitudes that should be used. In this paper, the minimum range of all sensor types of all sensor positions was used as common range. Thus, the common accelerometer range was set to $\pm 6g$ for all sensor positions. The common gyroscope range was set to $\pm 500^{\circ}/s$ for the sensor positions wrist, chest and hip and $\pm 1500^{\circ}/s$ for the ankle sensor. All amplitude values above those ranges were set to the minimum range value.



(a) Running



(b) Rope jumping

Figure 1: Energy signals of the accelerometer of the hip sensor. Examples are given for running and rope jumping.

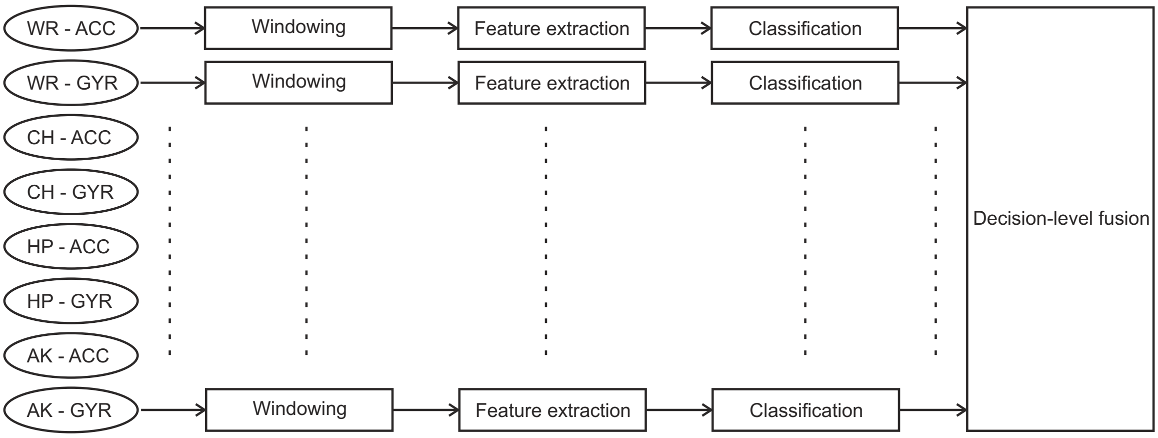


Figure 2: Proposed classification system. WR, CH, HP and AK define the sensor positions wrist, chest, hip and ankle, respectively. ACC and GYR define the sensor types accelerometer and gyroscope, respectively.

2.2.5 *Sampling Rate*

The fifth step was to define the common sampling rate. In this paper, the minimum available sampling rate was chosen, namely 100 Hz for all sensor types of all sensor positions.

2.2.6 *Data-Level Fusion*

The sixth step was to address the different orientations of the sensors in all available databases. In this paper, data-level fusion [4] was performed for each sensor type of each sensor position. Therefore, the energy signal $E[x]$ of the three axes a_1, a_2, a_3 of each sensor type was computed:

$$E[x] = \sum_{i=1}^{3} a_i^2[x]. \tag{1}$$

Fig. 1 shows the energy signals of the accelerometer of the hip sensor for two example activities.

2.3 Proposed Classification System

The proposed classification system is depicted in Fig. 2. For each available sensor type (accelerometer, gyroscope), windowing, feature extraction and classification were performed separately. In the final step, the classifier decisions of all sensor types were fused and a single activity was predicted. The system architecture was similar to [17]. The main difference was the additional splitting of the sensor position with respect to the sensor types. In the following section, the details of windowing, feature extraction and classification are described.

2.3.1 *Windowing*

Further processing of the acquired inertial data of the sensor types was performed in sliding windows. The width of the window was set to five seconds, which was also proposed in [10] and [17].

2.3.2 *Feature Extraction*

Nine features were computed for each sliding window in the energy signal. In order to extract information about the range of the signal amplitudes, the minimum and maximum of the amplitudes were computed. In order to extract information about the statistics of the signal amplitudes, the mean, variance, skewness and kurtosis of the amplitudes were computed. In order to extract information about the uncertainty of the signal, the normalized information entropy in the time domain was computed according to [11]. In order to extract frequency information of the signal, the spectral centroid and the bandwidth were computed according to [1].

2.3.3 *Classification*

For the classification of the eight selected daily life activities the Support Vector Machine (SVM) was used with a linear kernel [3]. For performance assessment, the mean class-dependent classification rate and the overall mean classification rate were computed based on a leave-one-subject-out cross-validation.

Table 3: Classification rates (in percent) regarding different sensor types. WR, CH, HP and AK define the sensor positions wrist, chest, hip and ankle, respectively. ACC and GYR define the sensor types accelerometer and gyroscope, respectively. The abbreviations for the activities are used according to Table 1. Classification rates are ranked in a descending order regarding the overall mean classification rates.

Rank ID	Sensor type	ST	WK	CS	RU	JP	VC	BC	RJ	Mean
1	**AK - GYR**	96.3	93.3	67.0	72.4	-	85.9	91.3	70.3	82.4
2	**HP - ACC**	98.2	87.1	48.9	86.7	84.7	63.9	92.3	84.6	80.8
3	**AK - ACC**	96.3	88.3	45.1	68.9	-	86.1	97.1	76.0	79.7
4	**WR - ACC**	78.9	89.4	43.1	88.5	-	64.3	79.7	94.4	76.9
5	**CH - ACC**	96.2	90.8	54.7	86.5	-	32.0	81.7	95.1	76.7
6	**HP - GYR**	98.3	78.3	41.1	74.6	47.9	22.4	71.3	77.5	63.9
7	**CH - GYR**	90.6	80.3	31.3	65.3	-	29.1	27.8	85.1	58.5
8	**WR - GYR**	85.9	83.5	9.4	32.2	-	45.8	73.8	68.8	57.1

The cost parameter of the SVM $C \in \{0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000\}$ was optimized by grid search and an inner leave-one-subject-out cross-validation loop. For the wrist, chest and ankle positions 28 subjects and for the hip position 33 subjects were available (Table 2). The sensor types were ranked in descending order according to the achieved overall mean classification rates.

2.3.4 *Decision-Level Fusion*

The classifier decisions of the sensor types were fused and a majority voting scheme was applied to obtain the final activity. In the case of an equal distribution of the predicted classes in the majority vote, the predicted class of the sensor type was chosen which was ranked higher. The performance of the decision-level fusion was determined for each database separately.

3. RESULTS

Table 3 shows the mean class-dependent classification rates and the overall mean classification rates regarding each sensor type. The best classification rate of 82.4 % was achieved by the gyroscope of the ankle sensor.

Regarding the three databases PAMAP2, DaLiAc and USC-HAD, the mean classification rates after decision-level fusion were 87.6 %, 87.6 % and 82.2 %, respectively. Thus, the overall mean classification rate averaged regarding all databases was 85.8 %.

4. DISCUSSION

Using big data will provide an immense advantage in future applications in activity recognition. This paper addressed the two attributes volume and variety of big data. A large database containing different sensor positions, sensor types, types of activities and execution of activities is needed in order to build up robust and stable activity recognition systems. Combining multiple databases is a first step to enlarge the database that is used for evaluation of algorithms. In this paper, the necessary steps were described which were needed to perform database fusion. Furthermore, an algorithm was developed which can deal with multiple databases varying in different number of sensor types and activities. In the following two sections, the proposed classification system and the database fusion are discussed.

4.1 Proposed Classification System

The best classification rate and the worst classification rate were achieved by the gyroscope of the ankle position and the gyroscope of the wrist position, respectively (Table 3). The reason might be that most of the considered activities included rotational movements of the lower extremities. The second best classification rate was achieved by the accelerometer of the hip position. The hip position is also preferred in literature since sensors near the body's center of mass cover a wide range of basic activities [10].

As a proof of concept, only the SVM with the described feature set was applied, since the main focus was the database fusion strategy. Further classifiers (e.g. RandomForest, Naive Bayes and kNN [19]) with additional features (e.g. wavelet and correlation between axes) will be applied in the future, which can easily be integrated in the proposed classification system.

Compared to the individual database classification results of 89 % in [13], 96 % in [23] and 89 % in [10], the proposed classification system achieved a lower classification rate of 85.8 % averaged regarding all databases. The reason might be the information loss due to the database fusion steps normalization and data-level fusion in this paper. Nevertheless, database fusion ensures more generalized results. In this paper, three databases were combined resulting in a total number of subjects of 42.

All in all, the findings in this paper showed the general applicability of the proposed classification system for combined databases.

4.2 Database Fusion

In the following section, the necessary steps for the combination of multiple databases are discussed.

The performance of the proposed classification system is dependent on the activities that should be classified. A further distinction between sitting, lying and standing as well as ascending and descending stairs requires additional information of the orientation of the sensors.

The performance of the proposed classification system is dependent on the used sensor types. In this paper, only accelerometer and gyroscope were used. In order to improve the performance of e.g. the distinction of walking and running, an additional heart rate monitor might help. The heart rate likely increases during high intensity running. The prediction of the additional sensor type can easily be integrated in the majority voting of the proposed classification system.

The unit adjustment and the normalization of the amplitudes were necessary because of the used feature types. Many features (e.g., the mean, minimum and maximum) have the same unit as the corresponding amplitudes. By using features that do not have a unit (e.g., correlation between two axes) these steps might be neglected.

In this paper, the lowest sampling rate regarding all three databases was used. Thus, the sampling rate was one bottleneck of the proposed classification system since the higher sampling rate in the DaLiAc database was not exploited. In the future, further research should focus on the influence of the sampling rate on the classification performance.

The performance of the proposed classification is dependent on the data-level fusion that was performed in this paper. By using the energy signal of each sensor type (Eq. 1), information about the orientation of the sensor axes was removed. In the future, further algorithms for data-level fusion should be considered, e.g. Kalman filtering [6].

In [5, 2, 14], big data was generated using their own study reaching a high number of subjects [5], a high variety of sensor orientations [2], and a high variety of sensor types [14]. Besides using only the own acquired data, researchers should consider database fusion strategies to further increase the amount of data and to exploit the advantages of different databases.

All in all, the paper listed the main steps that are required for database fusion. Database fusion is one strategy in order to generate a large amount of diverse data. High volume and high variety are two important attributes for big data. Using big data in the training of pattern recognition techniques will result in robust and stable activity recognition algorithms.

5. CONCLUSION

Activity recognition provides important feedback about the human's behavior. For the development of robust and stable activity recognition algorithms, there is a major need for databases that contain a large number of subjects, activities and sensors. In this paper, the strategy of database fusion was introduced, which is an important step towards big data for activity recognition. Furthermore, a classification system was proposed, which can deal with multiple databases varying in the number of sensors and activities. The proposed approach performed decision-level fusion and achieved an overall classification rate of 85.8 %.

In the future, it is possible to add physiological sensors to the system, e.g. heart rate monitor, to consider the axis orientation and to integrate additional databases. Pattern recognition systems, trained on big data, will provide reliable feedback about the individual quality of life, which results in motivating physically inactive people to be more active. This can lead to higher levels of health-related fitness, which should be one major goal for the future.

6. ACKNOWLEDGMENTS

This work was supported by the Bavarian Ministry for Economic Affairs, Infrastructure, Transport and Technology and the European fund for regional development.

7. REFERENCES

[1] G. Agostini, M. Longari, and E. Pollastri. Musical instrument timbres classification with spectral features. *EURASIP J Adv Sig Pr.*, 2003:5–14, 2003.

[2] O. Banos, M. Damas, H. Pomares, I. Rojas, M. A. Toth, and O. Amft. A benchmark dataset to evaluate sensor displacement in activity recognition. In *Proc UbiComp 2012*, pages 1026–1035, 2012.

[3] R. Duda, P. Hart, and D. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, New York, 2000.

[4] D. Hall and J. Llinas. An introduction to multisensor data fusion. *Proc. IEEE*, 85(1):6–23, 1997.

[5] Y. Hattori, S. Inoue, T. Masaki, G. Hirakawa, and O. Sudo. Gathering large scale human activity information using mobile sensor devices. In *Proc BWCCA 2010*, pages 708–713, 2010.

[6] R. E. Kalman. A new approach to linear filtering and prediction problems. *J Basic Eng-T ASME*, 82 (series D):35–45, 1960.

[7] N. Kawaguchi, N. Ogawa, Y. Iwasaki, K. Kaji, T. Terada, K. Murao, S. Inoue, Y. Kawahara, Y. Sumi, N. Nishio, and N. Ward. Hasc challenge: Gathering large scale human activity corpus for the real-world activity understandings. In *Proc AH 2011*, number 27, 2011.

[8] N. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell. A survey of mobile phone sensing. *IEEE Commun Mag*, 48(9):140–150, 2010.

[9] O. D. Lara and M. A. Labrador. A survey on human activity recognition using wearable sensors. *IEEE Commun Surv Tutor*, 15(3):1192–1209, 2013.

[10] H. Leutheuser, D. Schuldhaus, and B. M. Eskofier. Hierarchical, multi-sensor based classification of daily life activities: Comparison with state-of-the-art algorithms using a benchmark dataset. *PLOS ONE*, 8(10):e75196, 2013.

[11] A. Paraschiv-Ionescu, C. Perruchoud, E. Buchser, and K. Aminian. Barcoding human physical activity to assess chronic pain conditions. *PLOS ONE*, 7(2):e32239, 2012.

[12] Physical Activity Guidelines Advisory Committee. Physical activity guidelines advisory committee report, 2008.

[13] A. Reiss and D. Stricker. Creating and benchmarking a new dataset for physical activity monitoring. In *Proc PETRA 2012*, no pagination, 2012.

[14] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha, J. Doppler, C. Holzmann, M. Kurz, G. Holl, H. Chavarriaga, R. Sagha, H. Bayati, M. Creatura, and J. d. R. Millàn. Collecting complex activity datasets in highly rich networked sensor environments. In *Proc INSS 2010*, pages 233–240, 2010.

[15] P. Russom. Big data analytics. In *TDWI Best Practices Report, Fourth Quarter*, 2011, Accessed on May 30th 2014, http://tdwi.org/bdr-rpt.aspx.

[16] J. F. Sallis and B. E. Saelens. Assessment of physical activity by self-report: status, limitations, and future directions. *Res Q Exerc Sport*, 71 (supplement 2):1–14, 2000.

[17] D. Schuldhaus, H. Leutheuser, and B. M. Eskofier. Classification of daily life activities by decision level fusion of inertial sensor data. In *Proc BodyNets 2013*, pages 77–82, 2013.

[18] R. J. Shephard. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med*, 37(3):197–206, 2003.
[19] S. Theodoridis and K. Koutroumbas. *Pattern Recognition (4th Edition)*. Academic Press, San Diego, 2008.
[20] J. M. Warren, U. Ekelund, H. Besson, A. Mezzani, N. Geladas, and L. Vanhees. 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. *Eur J Cardiovasc Prev Rehabil*, 17(2):127–139, 2010.
[21] World Health Organization. Prevalence of insufficient physical activity. Accessed on May 30th 2014, http://www.who.int.
[22] M. Zhang and A. A. Sawchuk. Usc-had: A daily activity dataset for ubiquitous activity recognition using wearable sensors. In *Proc UbiComp 2012*, pages 1036–1043, 2012.
[23] M. Zhang and A. A. Sawchuk. Human daily activity recognition with sparse representation using wearable sensors. *IEEE J Biomed Health Inform*, 17(3):553–560, 2013.