

Comparative Evaluation of Feature Extraction Methods for Human Motion Detection

Olga Politi¹, Iosif Mporas², Vasileios Megalooikonomou¹

Multidimensional Data Analysis and Knowledge Management Laboratory
Dept. of Computer Engineering and Informatics, University of Patras
26500 Rion-Patras, Greece

¹{politi, vasilis}@ceid.upatras.gr, ²imporas@upatras.gr

Abstract. In this article we conduct an evaluation of feature extraction methods for the problem of human motion detection based on 3-dimensional inertial sensor data. For the purpose of this study, different preprocessing methods are used, and statistical as well as physical features are extracted from the motion signals. At each step, state-of-the-art methods are applied, and the produced results are finally compared in order to evaluate the importance of the applied feature extraction and preprocessing combinations, for the human activity recognition task.

Keywords: Accelerometers, movement classification, human motion recognition.

1 Introduction

One of the most important tasks in pervasive computing is to provide accurate and opportune information on people's activities and behaviors. Applications in medicine, security and entertainment constitute examples of this effort. For instance, patients with obesity, diabetes or heart disease, are often required to fulfill a program of physical exercise that is integrated within their daily activities [1]. In computer vision, complex sensors such as cameras have been used to recognize human activities but their accuracy falls under a real-home setting, due to the high-level activities that take place in the natural environments, as well as the variable lighting or clutter [8]. As a result, body-attached accelerometers are commonly used as an alternative in order to assess variable daily living activities.

The human motion detection problem using accelerometers is an emerging area of research. Sensors embedded in objects or attached on the body are generally chosen to study movement patterns or human behavior. Accelerometers have been used extensively due to their low-power requirements, small size, non-intrusiveness and ability to provide data regarding human motion. In a common scenario, these data can be processed using signal processing and pattern recognition methods, in order to obtain a real-time recognition of human motion.

Several human activity recognition systems have been proposed in the past, which include the use of accelerometers. Some of them analyze and classify different kinds of activity using acceleration signals [2], [3], while others apply them for recognizing a wide set of daily physical activities [4], or describe a human activity recognition framework based on feature selection techniques [5]. Bernecker et al [6], proposed a reclassification step that increases accuracy of motion recognition. Karantonis et al. [7] introduced an on-board processing technique for a real-time classification system, yielding results that demonstrate the feasibility of implementing an accelerometer-based, real-time movement classifier using embedded intelligence.

Khan et al. [8] propose a system that uses a hierarchical recognition scheme, i.e., the state recognition at the lower level using statistical features and the activity recognition at the upper level using the augmented feature vector followed by linear discriminant analysis. Several powerful machine learning algorithms have been proposed in the literature for the detection of human motion. The most widely used are the artificial neural networks [6, 8, 10], the naïve-Bayes [4, 14] and the support vector machines [5, 14].

In this paper, a comparative evaluation of feature extraction methods for human motion detection is presented. The main contribution of the paper is the proposal of the best combination between preprocessing and feature extraction, for the human activity recognition task. For this purpose, we choose to evaluate different preprocessing and feature extraction combinations, formed after thoroughly examining the existing state-of-the-art methods. Although this study may seem quite simple, there is little known in bibliography regarding official comparison between methods that include preprocessing combined with feature extraction. Since most studies concentrate on preprocessing or feature extraction separately, focusing on the above combinations seems important not only for the acquisition of better results, but also for discovering meaningful data interpretations and features that “characterize” human motion.

The rest of this paper is organized as follows: In Section 2 we present the framework constructed for human motion detection. Section 3 offers details about the experimental setup and in Section 4 we present the achieved experimental results. Finally in Section 5 we conclude this work.

2 Framework for Comparative Evaluation

In the present framework for comparative evaluation, we assume that the input to the framework consists of 3-dimensional $\{x, y, z\}$ signal streams, as illustrated in Fig. 1. Each stream represents one movement direction in the sense of moving forward/backward, up/down and left/right. Preprocessing \mathcal{P}_i consists of applying a sliding window W to the incoming streams, of constant length, resulting to W_i frames, where $1 \leq i \leq I$. The time-shift between two successive frames is also constant and can result to frame sequences with or without overlap. After applying the sliding win-

dow, each signal frame W_i is either led directly to the feature extraction stage (Non-preprocessing method) or is previously processed using two different techniques.

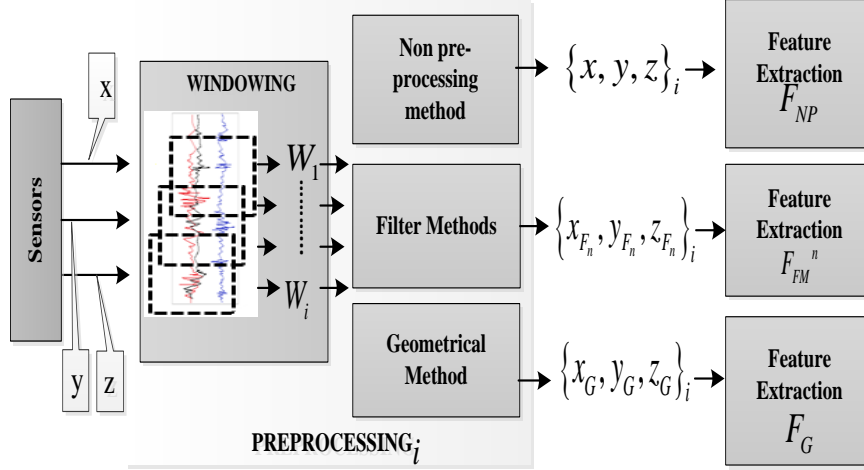


Fig. 1. Block diagram of the preprocessing stage of the motion detection framework

Specifically, the framework uses filtering methods F_n , with $1 \leq n \leq N$, that produce n signal frame representations per W_i frame and b) a geometrical representation method G . Each one of the previously mentioned methods yields different 3-dimensional data representations per frame W_i , namely $\{x, y, z\}$ for non-preprocessed data, $\{x_{F_n}, y_{F_n}, z_{F_n}\}$ for n chosen filters, $1 \leq n \leq N$, and $\{x_G, y_G, z_G\}$ for the geometrical representation. Each representation produced by the preprocessing stage is led to the feature extraction stage.

In this way, we have feature extraction carried out in parallel for the non-preprocessing representation, F_{NP} , for n filter methods representation, F_{FM}^n , and for the geometrical representation, F_G , per signal frame W_i . The block diagram of the feature extraction and classification stage of the motion detection framework is depicted in Fig. 2. In the feature extraction stage, every input is processed by statistical and physical feature extraction algorithms. The statistical algorithms can briefly be divided to time, frequency and time-frequency domain methods.

The utilization of these methods ensures that we get as much possible information from the data retrieved. In detail, each incoming representation of a signal frame is processed in parallel by each one of the feature extraction modules shown in Fig. 2. The estimated feature vectors could be mathematically represented as follows: the

time-domain features $F_T^i \in \mathbb{R}^{|F_T^i|}$, the frequency-domain features $F_F^i \in \mathbb{R}^{|F_F^i|}$, the time-frequency domain features $F_{TF}^i \in \mathbb{R}^{|F_{TF}^i|}$ and the physical features $F_P^i \in \mathbb{R}^{|F_P^i|}$.

After the decomposition of the different incoming frame representations to the described feature vectors, an individual classifier is build per feature vector. Specifically, in the classification stage, for time-domain features we have C_T classifier, for frequency domain features C_F classifier, for time-frequency C_{TF} and finally for physical features C_P classifier.

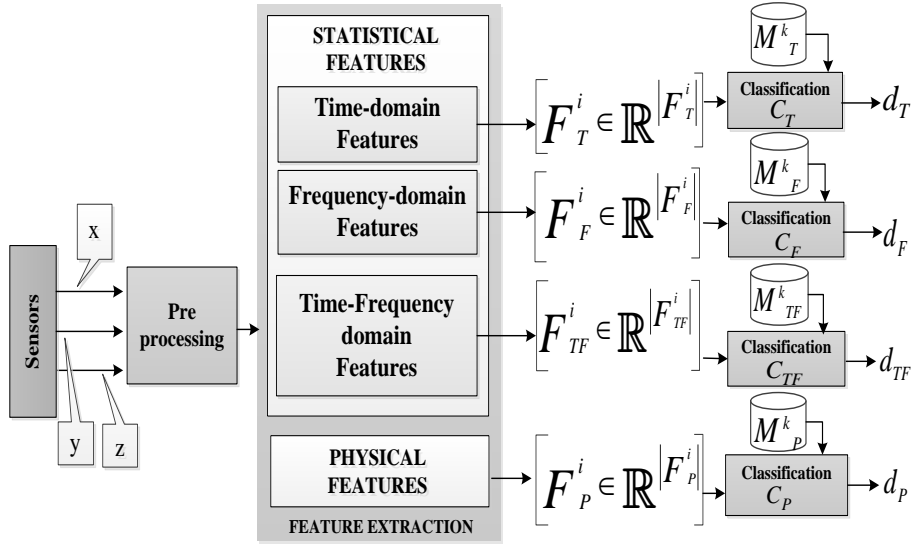


Fig. 2. Block diagram of the feature extraction and classification stage of the motion detection framework

During the training phase a set of motion data (training data) with known labels, i.e. with a-priori annotated motion labels, is used to estimate models, M_f^k , with $1 \leq k \leq K$ and $f = \{T, F, TF, P\}$, for each human motion k of interest and for each feature vector f . At the test phase the unknown motion data (test data) will be pre-processed and decomposed to feature vectors as in the training phase. The classification algorithms C_f will compare each vector F_f^j , with $f = \{T, F, TF, P\}$ and $1 \leq j \leq J$ against each motion model M_f^k in order to decide the corresponding motion class, i.e. $d_j = \arg \max_k \{C_f(F_f^j, M_f^k)\}$, where d_j is the motion class

label assigned to the j -th test frame of the sensor data. After classification a post-processing algorithm could be applied on the automatically labeled frames of the test recording in order to fine-tune the detected human motion classes, but this is left for further studying.

3 Experimental Setup

3.1 Sensor data description

The USC-HAD database [9], which is a freely available dataset provided by the University of Southern California was chosen in order to conduct experiments using the framework for comparative evaluation described in the previous section. The dataset corresponds to well-defined low-level daily activities appropriate for evaluation of algorithms for human motion in healthcare scenarios. The database consist of sensor measurements from 14 subjects, 7 male and 7 female, taken from a sensor placed at their front right hip, using sampling frequency equal to 100 Hz.

The activities performed were: walking forward/left/right/upstairs/downstairs, jumping, sitting, standing, sleeping, and elevator up/down. MotionNode, the sensing platform used, integrates a 3-axis accelerometer, a 3-axis gyroscope, and a 3-axis magnetometer. For the purpose of this paper, only the 3-axis accelerometer data were used. Subjects performed 14 different types of activities, including walking, running, sitting, standing, ascending and descending stairs. Ground truth was annotated by an observer while the experiments were being carried out.

3.2 Preprocessing and Feature Extraction

During preprocessing, each of the 3-dimensional $\{x, y, z\}$ signal streams was frame blocked with a sliding window of 1 second length, with time-shift 0.5 seconds [7]. Except the frame blocking three preprocessing methods were used [7], namely: (i) the gravitational acceleration (GA) component of the signal, (ii) the body acceleration (BA) component of the signal and (iii) the Tilt sensing, defined as the angular position with respect to gravity per axis. Together with the baseline preprocessing (non-preprocessing method), four preprocessing setups were created.

For each of the four preprocessed outputs per frame, statistical and physical features were extracted [5]. As far as feature analysis is concerned, statistical features are briefly divided into *time-domain*, *frequency domain* and *time-frequency domain* [11]. Time-domain features include mean, median, variance, root mean square (RMS), standard deviation, skewness, kurtosis and interquartile range (25%, 50%, 75%). Frequency-domain features, which mainly represent the periodic structure of the signal, are the Fourier transform, the spectral entropy, the spectral energy and the 3rd order autoregressive-filter (AR) coefficients.

Time-frequency domain features are used to investigate both time and frequency characteristics of complex signals and in general employ wavelet techniques, such as wavelet coefficients or energy of wavelet coefficients. The physical features [5] in-

clude the movement intensity (MI), the eigenvalues of dominant directions (EVA), correlation of acceleration along gravity and heading directions (CAGH) and the averaged velocity along each direction and the averaged acceleration energy (AAE).

3.3 Classification

For each of the four preprocessing methods the estimated feature vector has dimensionality equal to 77. Thus, the final feature vector per frame is of dimensionality $4 \times 77 = 308$. For the classification of the estimated feature vectors, we relied on the support vector machines (SVMs) implemented with the sequential minimal optimization method [13] using the polynomial kernel function (poly). The choice of the kernel-based algorithm is owed to the fact that SVMs do not suffer from the curse of dimensionality phenomenon [12]. The classifier was implemented using the WEKA machine learning toolkit software [13].

4 Experimental Results

The human motion detection framework presented in Section 2 was evaluated using the experimental setup described in Section 3. To avoid overlap between the training and test datasets we followed a 10-fold cross validation protocol. The experimental results, in percentages, for the four feature extraction methods and for each of the 14 subjects of the dataset are presented in Table 1. The best accuracy achieved for each subject is indicated in bold. The classification was carried out using individual classifiers, under a subject-dependent scenario.

Table 1. Motion detection accuracy (%) per subject and feature extraction method

Sub.	NP	BA	GA	Tilt
01	91,80	57,66	91,88	89,72
02	90,55	62,93	90,29	89,01
03	90,74	61,92	90,63	89,80
04	85,93	55,33	85,42	83,94
05	86,23	59,00	86,74	85,10
06	91,99	92,51	56,28	90,42
07	92,46	60,56	92,72	90,55
08	91,79	65,61	91,35	90,10
09	92,07	69,24	91,24	91,08
10	91,72	64,56	91,42	88,27
11	93,91	58,54	93,89	91,37
12	93,34	63,44	93,51	92,81
13	92,57	57,61	92,33	90,81
14	87,45	53,98	86,67	83,48

From Table 1 we can see that the best accuracy achieved is approximately 94%, in the non-preprocessed data representation and for Subject 11. A better view of the results is shown in Fig. 3, where it can be seen that the BA method is the one that performs the worst among feature extraction methods. Both the non preprocessed method and the tilt representation achieve very good results, with GA feature extraction method following shortly after the first two methods.

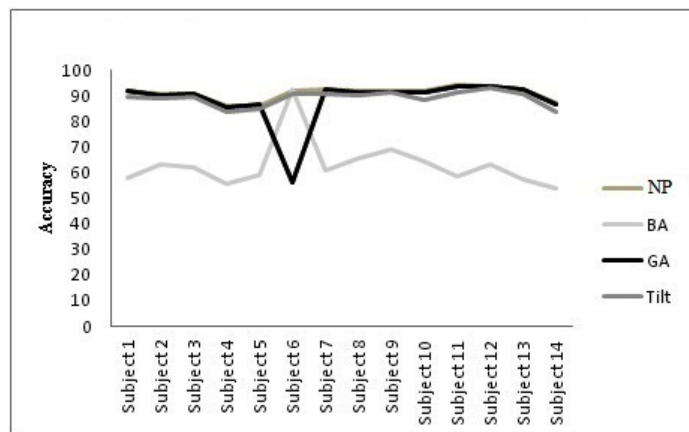


Fig. 3. Motion detection accuracy per subject and feature extraction method

The experimental results may show that the non preprocessed method performs the best, so not using preprocessing can yield good accuracy in motion detection, but the interesting finding is that the Tilt method also performs as well. This is very important, due to the fact that by using Tilt sensing, we achieve to add a geometrical aspect to the previously linear acceleration signals. As a result, when it comes to motions that are characterized by changes in angular velocity or torque, i.e. turning or clockwise movements that may be followed by movements that are linearly translated, the Tilt sensing method is preferable.

Another observation is that the GA method also performs well, apart from the case of Subject 6. This outcome may be due to the fact that the database used for the experiments includes many motions that have a strong relation to gravity acceleration. To be more specific, five out of fourteen activities (walking upstairs/downstairs, elevator up/down and jumping) could be translated to changes only in the gravity dimension, not to mention the three static activities (lying, standing and sitting) which are also “characterized” by gravity acceleration.

As a result, to face the problem of variability in human motion, it seems more appropriate to deal with data preprocessing that involves mathematical or geometrical representation, rather than focus on the idea of applying filter-based signal processing. In addition, we investigated the importance of features used in the overall experimental process, in terms of feature influence in the classification process. The feature

ranking was performed using the ReliefF algorithm [15], and the top-3 features ranked by ReliefF are presented in Table 2.

Table 2. Top-3 features ranked by the ReliefF algorithm

Ranking	Features
1	Spectral entropy
2	CAGH of x axis (gravity)
3	Correlation xz

As can be seen in Table 2, within the most discriminative motion features for the task of automatic motion recognition are spectral entropy, the correlation between acceleration along gravity direction (CAGH), and the correlation between gravity and heading direction. At this point, it is evident that frequency domain features play an important role, as we meet spectral entropy and correlation at the top three ranked features. A physical feature (CAGH) is second in ranking, hence the importance of the addition of physical features in the feature extraction process. Attempting to interpret these ranking results, we could initially observe that spectral entropy is justifiably ranked first, since it helps differentiate between signals that have similar energy values but correspond to different activity patterns [16]. The USC-HAD database includes movements with similar energy, i.e. sitting compared to standing and lying, or walking in different patterns (forward, left, right, upstairs and downstairs).

Following in rank is the CAGH feature that improves differentiation between activities that involve translation (in terms of acceleration intensity) in a single dimension, compared to the total acceleration intensity of the other two dimensions. Movements such as walking upstairs/ downstairs and walking left/right/forward fall in the previous category, hence the importance of using the CAGH feature. Correlation xz (between gravity and heading direction) also seems to play an important role in the classification results. Its interpretation is similar to CAGH, only in this case the comparison is inner-axial. Motions that have a strong xz correlation in the database are jumping, elevator up/down as well as static movements (lying, sitting, and standing).

Finally, we observe that statistical features are not present in the top-3 ranking, but this does not indicate that they are not important in the classification process, since the overall accuracy is not obtained only by the use of the top-3 features; it is rather achieved using the whole feature set. The reason we present only the top-3 features lies with the effort to see the most influential and perhaps most meaningful features, but those features alone cannot yield high accuracies for motion detection.

5 Conclusion

In the present work we present a comparative evaluation of feature extraction methods, based on 3-dimensional data acquired from inertial sensors worn by humans. The framework uses four different preprocessing methods and the motion signals are parameterized by statistical and physical features. The experimental results indicated an average performance of 90.89% for the best combination of preprocessing and

feature extraction methods, for 14 everyday human motion activities. The highest performance achieved was approximately 94%. We deem the application of experiments constructed under a subject-independent scenario, in order to see if the present study responds well in intra-class errors, namely the errors introduced in the classification process when different people execute the same movement.

6 Acknowledgement

The reported research was partially supported by the BioMedMine project (Thalis – University Of Patras – Mining Biomedical Data And Images: Development Of Algorithms and Applications), as well as the ARMOR Project (FP7-ICT-2011-5.1 - 287720) “Advanced multi-parametric Monitoring and analysis for diagnosis and Optimal management of epilepsy and Related brain disorders”.

7 References

1. Jia, Y.: Diabetic and exercise therapy against diabetes mellitus. In: Second International Conference on Intelligent Networks and Intelligent Systems, pp. 693–696 (2009)
2. Mantyjarvi, J., Himberg, J., Seppanen, T.: Recognizing human motion with multiple acceleration sensors. In: IEEE Int. Conf. Syst., Man, Cybern., vol. 2, pp. 747–752, (2001)
3. Sekine, M., Tamura, T., Akay, M., Fujimoto, T., Togawa, T., Fukui, Y.: Discrimination of walking patterns using wavelet-based fractal analysis. In: IEEE Trans. Neural Syst. Rehabil. Eng., vol. 10, no. 3, pp. 188–196, Sep. (2002)
4. Ermes, M., Parkka, J., Mantyjarvi, J., Korhonen, I.: Frequent Bit Pattern Mining Over Triaxial Accelerometer Data Streams for Recognizing Human Activities and Detecting Fall. In: Procedia Computer Science, Volume 19, Pages 56–63 (2013)
5. Zhang, M., Sawchuk, A.: A Feature Selection-Based Framework for Human Activity Recognition Using Wearable Multimodal Sensors. In: BodyNets 2011, University of Southern California, Los Angeles, (2011)
6. Bernecker, T., Graf, F., Kriegel, H., Moennig, C.: Activity Recognition on 3D Accelerometer Data. Technical Report, (2012)
7. Karantonis, D.M., Narayanan, M.R., Mathie, M., Lovell, N.H., Celler, B.G.: Implementation of a Real-Time Human Movement Classifier Using a Triaxial Accelerometer for Ambulatory Monitoring. In: IEEE on information technology in biomedicine, vol. 10, (2006)
8. Khan, A., Lee, Y., Lee, S.Y., Kim, T.: Triaxial Accelerometer-Based Physical-Activity Recognition via Augmented-Signal Features and a Hierarchical Recognizer. In: IEEE transactions on information technology in biomedicine, vol. 14, no. 5 (2010)
9. Zhang, M., Sawchuk, A.: USC-HAD: A Daily Activity Dataset for Ubiquitous Activity Recognition Using Wearable Sensors. In: UbiComp '12, USA (2012)
10. Khan, A. M., Lee, Y. K., Lee, S.Y.: Accelerometer's Position Free Human Activity Recognition Using A Hierarchical Recognition Model. In: IEEE HealthCom (2010)
11. Avci, A., Bosch, S., Marin-Perianu, M., Marin-Perianu, R., Havinga, P.: Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey. University of Twente, The Netherlands (2010)

12. Burges C., A tutorial on Support Vector Machines for Pattern Recognition. In: Data Mining and Knowledge Discovery, Vol. 2(2), p.121-167, Kluwer Academic Publishers, (1998)
13. Keerthi, S.S., Shevade, S.K., Bhattacharyya, C., Murthy K.R.K: Improvements to Platt's SMO Algorithm for SVM Classifier Design. Neural Computation. 13(3):637-649 (2001)
14. Ravi, N., Dandekar, N., Mysore , P., Littman, M.L.: Activity Recognition from Accelerometer Data, Department of Computer Science, Rutgers University (2005)
15. Robnik-Sikonja, M., Kononenko, I.: An adaptation of Relief for attribute estimation in regression, ICML '97, 1997, pp. 296–304.
16. Brezmes, T., Gorricho, J.L., and Cotrina, J.: Activity recognition from accelerometer data on a mobile phone,” Soft Computing and Ambient Assisted Living, 2009, vol. 5518, pp. 796–79.