The 11th International Conference on Mobile Systems and Pervasive Computing (MobiSPC-2014)

A Study on Human Activity Recognition Using Accelerometer Data from Smartphones

Akram Bayat*, Marc Pomplun, Duc A. Tran

Department of Computer Science, University of Massachusetts, Boston, 100 Morrissey Blvd Boston, MA 02125, USA

Abstract

This paper describes how to recognize certain types of human physical activities using acceleration data generated by a user's cell phone. We propose a recognition system in which a new digital low-pass filter is designed in order to isolate the component of gravity acceleration from that of body acceleration in the raw data. The system was trained and tested in an experiment with multiple human subjects in real-world conditions. Several classifiers were tested using various statistical features. High-frequency and low-frequency components of the data were taken into account. We selected five classifiers each offering good performance for recognizing our set of activities and investigated how to combine them into an optimal set of classifiers. We found that using the average of probabilities as the fusion method could reach an overall accuracy rate of 91.15%.
ⓒ 2014 The Authors. Published by Elsevier B.V.
Selection and peer-review under responsibility of Elhadi M. Shakshuki.

Keywords: Activity Recognition, Smartphone, Accelerometer

1. Introduction

Human activity recognition is an important yet challenging research area with many applications in healthcare, smart environments, and homeland security[4,15]. Computer vision-based techniques have widely been used for human activity tracking, but they mostly require infrastructure support, for example, installation of video cameras in the monitoring areas[16]. Alternatively, a more efficient approach is to process the data from inertial measurement unit sensors worn on a user's body (e.g.,[3,13]) or built in a user's smartphone (e.g.,[18,6]) to track his or her motion.

We aim to develop a model that is capable of recognizing multiple sets of daily activities under real-world conditions, using data collected by a single triaxial accelerometer built into a cell phone (in our study, an Android smartphone). A triaxial accelerometer is a sensor that returns an estimate of acceleration along the x, y and z axes from which velocity and displacement can also be estimated. Activity recognition is formulated as a supervised classification problem, whose training data is obtained via an experiment having human subjects perform each of the activities.

* Corresponding author. Tel.: +0-617-287-6485 ; fax: +0-617-287-6433.
   E-mail address: akram@cs.umb.edu

Our objective differs from earlier studies as follows. First, some of the activities in our study have not been widely studied (e.g., slow versus fast walking, aerobic dancing). Second, we want to use less sensory input than existing work requires, yet able to obtain a comparable accuracy. Last but not least, we aim at a classification methodology that is robust regardlessly of the classifier tool in use.

Our model has been tested in an experiment having four users each performing one of the following six physical activity patterns: Slow Walking, Fast Walking, Running, Stairs-Up, Stairs-Down, and (Aerobic) Dancing. In the following sections, we discuss the related work, describe our data collection methodology and our approach to recognize activity from accelerometer data, and results of our experiment.

2. Related Work

Some of the activity recognition works focus on the use of multiple accelerometers and possibly other sensors. Lester et al.[5] developed an automatic physical activities recognition system in a controlled environment using accelerometers and microphones. A wavelet-based activity classification method using one or more accelerometers was proposed in[8] by Mannini and Sabitini, in which the dynamic motion component is separated from the gravity components, leading to a 98.4% of accuracy. Foerster and Fahrenberg[14] collected data from 31 subjects using five accelerometers and built a hierarchical classification model in order to identify different body postures and movements.

Wearable devices can also be used to monitor human activities. Casale et al.[3] use a wearable device for collecting acceleration data for human activity recognition, obtaining a 94% of accuracy. In the work of Bao and Intille[2], five biaxial accelerometers are worn on the user's right hip, dominant wrist, non-dominant upper arm, dominant ankle, and non-dominant thigh to monitor 20 types of activities, trained with 20 users using decision tables, instance-based learning, C4.5 and Naive Bayes classifiers. Their results indicate that the accelerometer placed on the thigh was the most powerful for distinguishing between activities. Krishnan et. al.[13] collected data from ten subjects using three wearable accelerometers to detect lower body motions. In[7] by Nishkam and Nikhil, a single triaxial accelerometer is worn near the pelvic region to recognize eight activities: Standing, Walking, Running, Stairs-Up, Stairs-Down, Sit-Ups, Vacuuming, Brushing Teeth. Four different settings are considered for the dataset: (1) data collected for a single subject over different days, (2) data collected for multiple subjects over different days, (3) data collected for a single subject on one day as training data and further data collected for the same subject on another day as testing data, and (4) data collected for a subject for one day as training data and more data on another subject on another day as testing data. Their experimental results find the Plurality Voting classifier the best for the first three settings. It is also found difficult to distinguish Stairs-Up from Stairs-Down or Running, and impossible to distinguish Teeth Brushing versus Stairs-Down activities. Their classification accuracy for the forth setting is 73.33%.

The use of a smartphone's acceleration data is investigated in[18,6]. Data was collected from twenty-nine users, each carrying an Android phone in their pocket as they performed six activities: Walking, Jogging, Stairs-Up, Stairs-Down, Sitting, and Standing. Three learning algorithms were evaluated: Logistic Regression, J48, and Multilayer Perceptron. According to their results, the overall accuracy is above 90%, except for the case Stairs-Up versus Stairs-Down that poses greater difficulty. They also found that neural network classification can best detect Jogging and Stairs-Up while J48 can effectively detect the other activities.

3. System Design

In this section we describe step-by-step how to discriminate between different activities using the acceleration data collected from a smartphone. First, we need to collect the data. Then we analyze the data to extract informative features. Finally, we use these data for training to build a model based on the selected features and validate it.

3.1. Data Collection

In order to collect data, we used a triaxial accelerometer in the Android phone to measure acceleration. Data from this accelerometer includes the acceleration along the x-axis, y-axis and z-axis. These axes capture the horizontal/sideway movement of the user (x-axis), upward/downward movement (y-axis), and forward/backward movement (z-axis). Figure 1(a) demonstrates these axes relative to a user.

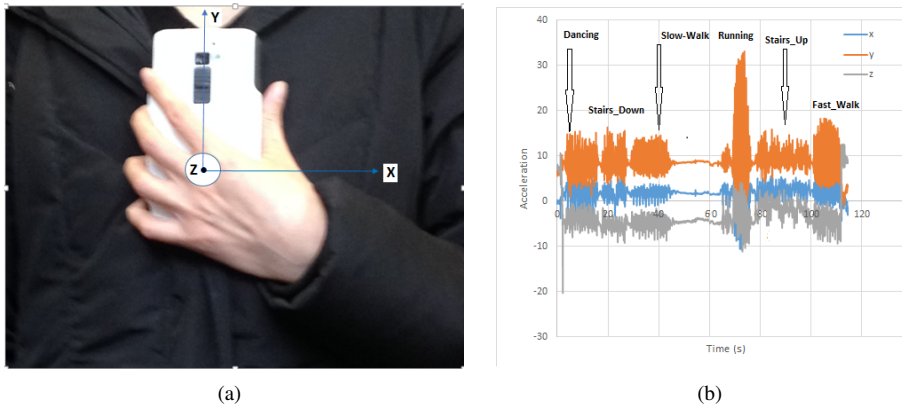(a)   (b)

Fig. 1. (left): Android smartphone with a built-in accelerometer and its axis directions; (right): A subject's triaxial accelerometer data over time in different activities.
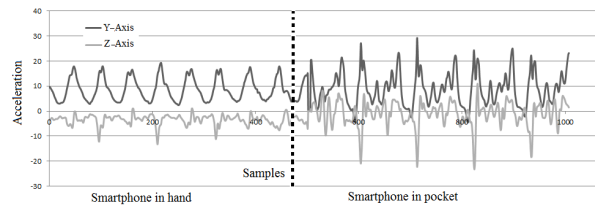

Fig. 2. Sample acceleration signals for fast walking (performed by an identical subject) with the smartphone in two different positions.

Four subjects, two male and two female between 29 to 33 years of age, volunteered to participate in this research study. These subjects, each carrying a cell phone, performed all the six activities: (1) Running: The user is running, (2) Slow-Walk: The user is walking with slow pace, (3) Fast-Walk: The user is walking with fast pace, (4) Aerobic Dancing: The user is moving rhythmically, (5) Stairs-Up: The user is going up a staircase, and (6) Stairs-Down: The user is going down a staircase.

No noise cancellation was done on the data. To avoid mislabeling, subjects were asked to stop and wait a few seconds after an activity before starting the next activity. Acceleration data collected between the start and stop times of an activity were labeled with the name of that activity. The sampling rate was 100Hz; i.e., one sample for every 10ms. Each subject provided between 180 and 280 seconds of data. Overall, 79,573 samples of three-dimensional acceleration data were obtained. Figure 1(b) shows an example of triaxial accelerometer data for a user performing different activities.

During the collecting of the data, one may ask where on the subject's body we should place the smartphone. This question arises since the acceleration data varies widely for the same physical activity at different positions on a user's body. This is clearly illustrated in Figure 2 which shows the acceleration data of fast walking for a subject when he holds the phone in his hand versus in his pocket. In our experiment, therefore, a subject cannot change the position of the smartphone (hand to pocket, pocket to hand, etc.) during the data collection process. Otherwise, the acceleration data can change significantly for the same activity, resulting in inaccurate recognition.

3.2. Feature Evaluation

Feature creation is a critical step in the development of any classifier. An activity recognition system does not solve the classification task directly on raw acceleration data. Generally, the classification is performed after an informative data representation is created in terms of feature vectors.

In our work, the accelerometer generates three time series along x-axis, y-axis and z-axis, denoted by A_X, A_Y and A_Z, respectively. Each time series combines the linear acceleration due to body motion and due to gravity. We propose a digital low pass filter in order to separate the AC component from the DC component in each time series. The high-
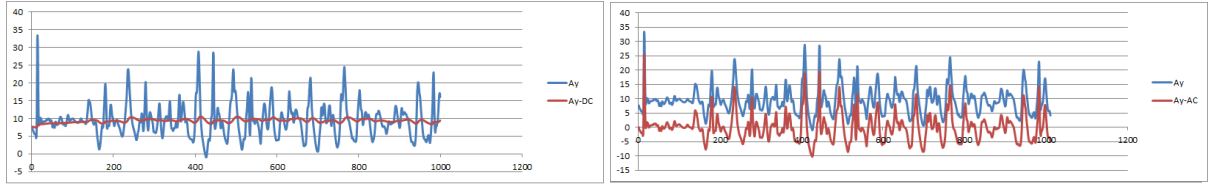
Fig. 3. Low pass filter (top figure) results in a signal containing only the gravitational acceleration, high pass filter (bottom figure) results in a signal containing only the body acceleration.

Fig. 4. Acceleration plots for the six activities along the z-axis that captures the forward movements. All six activities exhibit periodic behavior but have distinctive pattern. Running and Fast-Walking exhibit similar patterns.

frequency component, the AC component, is mostly related to the dynamic motion the subject is performing such as walking or running. On the other hand, the low-frequency component of the acceleration signal, the DC component, is mainly tied to the influence of gravity [5,10]. Setting the cut-off frequency to 0.25Hz, our low pass filter is calculated as follows: $A_{DC}[n] = a_1 A[n] + b_1 A_{DC}[n-1]$, where A_{DC} is the filtered output data and A is the raw input data (A_X, A_Y, A_Z). The filter coefficients a_1 and b_1 are constants that are computed using sampling rate and cutoff frequency. The optimal cutoff frequency in order to exclude gravity component alone would range from 0.1 to 0.5 HZ[12].

Considering the cutoff frequency around 0.25 Hz and sampling rate of 100 samples per second, we find the following solution for the coefficients of the low pass filter:$a_1 = .0158, b_1 = .9843$. In order to find the body acceleration components, we subtract the low pass filtered data from original signal along three axes. The effect of low pass and high pass filtering on each time series is to create gravitational and body acceleration data in different directions. Figure 3 shows the effectiveness of the simple recursive lowpass filter for isolating the gravitational acceleration and body acceleration.

After isolating the DC and AC acceleration components, for each time series A_i ($i \in \{X, Y, Z\}$), two more time series A_{DC_i} and A_{AC_i} are obtained, respectively representing the DC component and AC component along each axis. We compute the magnitude of acceleration, A_m, across all directions as another time series as well. However, since we are only interested in body motion acceleration, we pick AC components and magnitude of acceleration to be used for feature construction.

3.3. Feature Extraction

To extract descriptive features for the four time series obtained in the previous section, we apply a technique of window overlapping[3]. Figure 4 shows the accelerometer data along the z-axis for each of the six activities. All activities exhibit periodic behaviors but with distinctive patterns. We can take advantage of their periodicity to calculate features
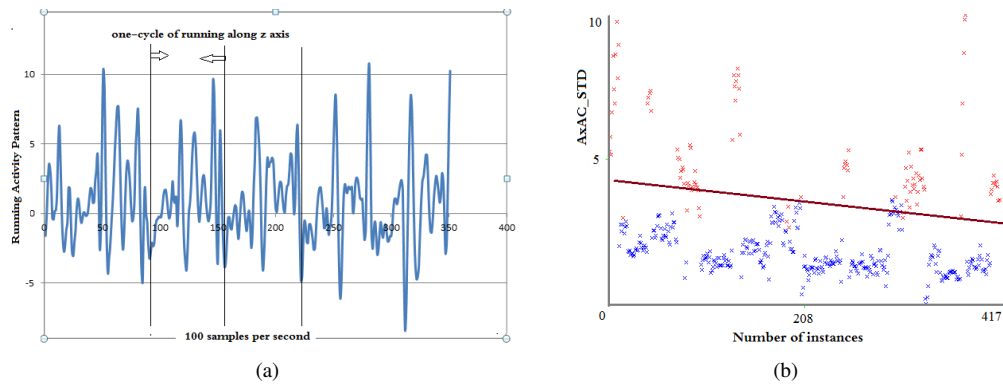
Fig. 5. (left): A few cycles of acceleration data along z-axis (forward) for Running; (right): AxAC_STD is the standard deviation of the AC component of acceleration along x-axis. This feature can distinguish the two clusters.

from consecutive windows with certain size. In window overlapping, we subdivide the data set into smaller subsets and window them individually. We choose a window of 128 samples, corresponding to 1.28 seconds of accelerometer data for feature extraction from each of the time series because it can sufficiently capture cycles in activities such as running, walking, or dancing (see Figure 5(a)). To reduce information loss at the edges of the window, the individual sets may overlap in time. For this purpose, features are computed with 64 samples overlapping, corresponding to 50% of overlap between consecutive windows.

The average on each axis over a given time period can serve as a good estimate for the gravity component. Therefore, we use the mean value for each window as one of the features. We also measure elapse time between consecutive local peaks in the periodic signals associated with each activity. Since average of time between signal peaks of each window is not quite useful for high intensity activities[18], we measure the average number of occurrences of peaks in each window instead of average time between peaks; we call this feature average of peak frequency (APF). In addition, the variance of APF along three axes is another feature that can measure how far a set of APF along x-axis, y-axis and z-axis is spread out; we name this feature "VarAPF". Based on our data observation for each activity, computing root mean square (RMS) value is especially useful since variants of activity signals can be positive or negative. Standard deviation is also calculated for each window to derive another feature as suggested in[1]. Furthermore, we propose using two more features: the Minmax value, which is defined as the difference between maximum and minimum for each window, and the correlation between different axes in order to improve the recognition of activities. In this way, we obtain a 24-dimensional feature vector.

In order to find the most informative features in our feature vector, we evaluate the features from the clustering point of view. Relevant features can discriminate clusters, while irrelevant features can not. For example, Figure 5(b) illustrates this notion where two clusters are created by utilizing knowledge from a feature which is defined as the standard deviation of the AC component of acceleration along the x-axis. Different subsets of relevant features may result in different clustering results, which greatly helps discover various hidden patterns in the data[11]. Therefore, we use a clustering method to select more relevant features in order to improve classification efficiency and quality. In the end, we select the following 18 relevant features from our original 24-dimensional feature vector: Mean along z-axis (1 feature), MinMax, STD and RMS for A_m (3 features), APF along x-axis, y-axis and z-axis (3), VarAPF (1), STD along x-axis, y-axis and z-axis (3), RMS along x-axis, y-axis and z-axis (3), Correlation between z-axis and y-axis (1), and MinMax along x-axis, y-axis and z-axis (3).

4. Classification

Six activities were studied as listed above. Subjects were requested to stop between two activities in order to note the start and end times of different activities. We used this information to label different activities for our supervised learning classification problem. Activity labels were chosen to reflect the content and style of the action. Two datasets

Table 1. Classification accuracy for different classifiers on two training datasets (phone in hand, phone in pocket).

Classifier	Accuracy (in-hand)	Accuracy (in-pocket)
Multilayer Perceptron	89.48%	89.72%
SVM	88.76%	72.27%
Random Forest	87.55%	85.15%
LMT	85.89%	85.04%
Simple Logistic	85.41%	85.05%
Logit Boost	82.54%	82.24%

Table 2. Classification accuracy as average of probabilities of different pools of classifiers (here, MP="Multilayer Perception").

IN-HAND		IN-POCKET	
Classifier	Accuracy	Classifier	Accuracy
MP, LogitBoost, SVM	91.15%	MP, Random Forest, SimpleLogistic	90.34%
MP, LogitBoost, SVM, LMT	90.90%	MP, LogitBoost, SimpleLogistic, Random Forest	90.34%
MP, LogitBoost, SVM, Random Forest	90.90%	MP, LogitBoost, SimpleLogistic	88.47%
MP, LogitBoost	88.51%	MP, Random ForesT, LMT, SimpleLogistic	88.16%
MP, SVM	88.27%	MP, LMT, SimpleLogistic	86.92%
MP, LogitBoost, SVM, Random Forest, LMT	81.10%	MP, LogitBoost, SimpleLogistic, Random Forest, LMT	83.18%

are considered, one for in-hand phone position and one for in-pocket phone position. Various supervised classification methods were performed.

4.1. Individual Classifiers

We evaluated the performance of the following classifiers, all available in the Weka toolkit: Multilayer Perceptron, Random Forest, LMT, SVM, Simple Logistic and LogitBoost. Classifiers were trained and tested using a 10-fold cross validation method on the set of extracted features.

The summary results for our activity recognition experiments are presented in Table 1 for two datasets: phone in hand position and phone in pocket position. Overall, Multilayer Perceptron offered the highest performance, yielding 89.48% accuracy for in-hand position and 89.72% accuracy for in-pocket position. SVM was the second most accurate for in-hand position. Its good performance has been supported by prior research on human activity recognition tasks. However, Our results also showed that Random Forest demonstrated high accuracy for both cases. Table 1 indicates that carrying the phone in pocket or in hand produced similar results. Hence, our recognition method is robust to smartphone position.

4.2. Combination of Classifiers

Combining multiple good classifiers can improve accuracy, efficiency and robustness over single classifiers[9]. The idea is that different classifiers may offer complementary and diverse information about patterns to be classified, allowing for potentially higher classification accuracy. We select the optimal set of classifiers from the previous section and then combine them by a fusion method. Two such fusion methods are considered: majority voting and average of probabilities. In the majority voting fusion method, each single classifier gives a vote for its prediction and the class receiving the most votes is the final prediction[10]. The average of probabilities fusion method returns the mean of the probability distributions for each of the single classifiers.

The fusion method we decided to use to combine the classifiers is average of probabilities as it was found in our experiment to offer better results than majority voting for our activity recognition model. Multiple different combinations are applied and our goal is to find the combination of classifiers with the highest accuracy using average of probabilities. Table 2 shows the classification accuracies for different combinations. For in-hand phone position, the combination consisting of Multilayer Perceptron, LogitBoost and SVM classifiers yields the highest accuracy of 91.15%. For in-pocket phone position, the best combination is of Multilayer Perceptron, Random Forest and
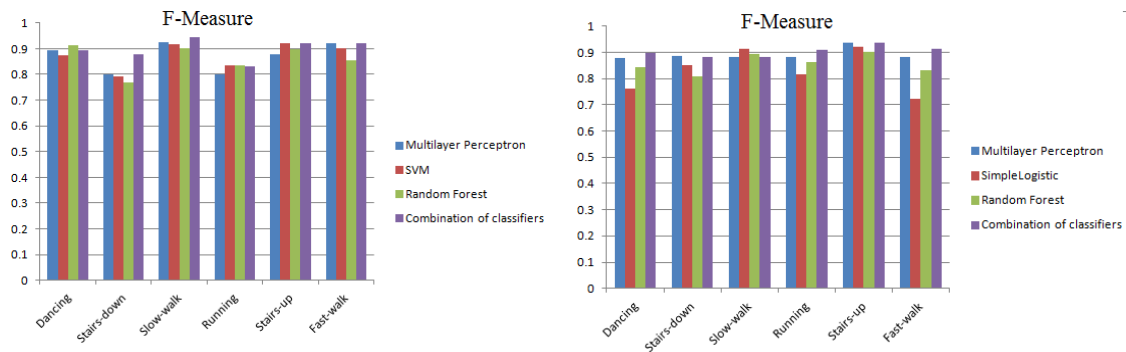
Fig. 6. F-measure for each activity using four classifiers: in-hand phone position (top figure) and in-pocket phone position (bottom figure).
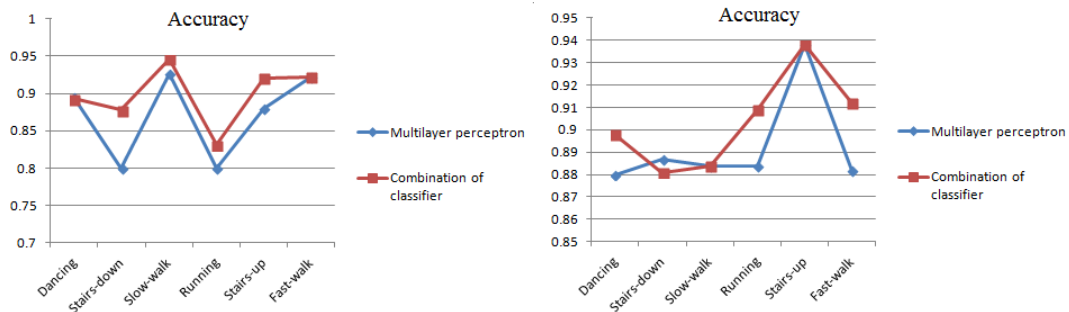

Fig. 7. F-measure comparison between best individual classifier versus best combination of classifiers: : in-hand phone position (top figure) and in-pocket phone position (bottom figure).

SimpleLogistic with 90.34% accuracy. Table 3 reports the Confusion Matrix associated with the best combination of classifiers for the two phone positions.

Table 3. Confusion Matrix for the combination of three best classifiers on two datasets: in-hand phone position and in-pocket phone position.

IN-HAND							IN-POCKET						
a	b	c	d	e	f	Classified as	a	b	c	d	e	f	Classified as
66	2	0	2	3	1	a=Dancing	54	1	3	1	1	0	a=Dancing
01	40	1	0	3	1	b=Stairs-Down	2	63	4	0	2	0	b=Stairs-Down
0	3	81	0	2	0	c=Slow-Walk	0	4	67	0	1	0	c=Slow-Walk
5	2	0	23	1	0	d=Running	1	0	1	24	0	1	d=Running
2	5	3	0	123	0	e=Stairs-Up	0	2	2	0	126	2	e=Stairs-Up
2	1	2	0	1	46	f=Fast-Walk	0	2	1	0	2	47	f=Fast-Walk

4.3. Individual vs. Combination

Figure 6 shows the F-measure of each activity using optimal individual classifiers and combination of classifiers. The F-measure considers both the precision and the recall of the classification. Good performance for each activity is obtained if that activity can be classified with high precision and recall. It is observed that the F-measure for Running and Dancing when the phone is held by hand is below 90%. However, the performance for detecting these two activities improves when subjects put the smartphone in their pants pocket (F-measure above 90%). Therefore, for the two highest-intensity activities, Running and Dancing, we achieve more accurate recognition when we use the data for in-pocket phone position. Figure 7 compares the performance of the best individual classifier versus the best combination of classifiers. The latter appears to be the better for both phone positions.

5. Conclusion

In this paper, recognition accuracy of up to 91.15% on various everyday activities using a single triaxial accelerometer was obtained. The data were acquired from multiple subjects under real-world conditions for two most common phone positions: smartphone in hand and smartphone in pants pocket. A new set of features was taken into account and different classifiers were used for evaluating recognition performance. Combining the three best classifiers using the average of probabilities method turned out to be the best classifier for activity recognition, outperforming all individual classifiers. We further showed that our recognition method can detect activities independent of smartphone's position. For future work, we plan to extend our activity recognition task in several ways. First, we intend to recognize additional activities, such as sleeping or bicycling. Second, we would like to collect data from more users of various ages. Third, we plan to extract more features that could better discriminate different activities.

Acknowledgments

References

1. Aminian, K and Robert, Ph and Buchser, EE and Rutschmann, B and Hayoz, D and Depairon, M Physical activity monitoring based on accelerometry: validation and comparison with video observation, Medical & biological engineering & computing, 1999; 37:3-304
2. Bao, Ling and Intille, Stephen S, Pervasive computing, Activity recognition from user-annotated acceleration data, 2004, Springer.
3. Casale, Pierluigi and Pujol, Oriol and Radeva, Petia, Human activity recognition from accelerometer data using a wearable device, Pattern Recognition and Image Analysis,2011;289, Springer.
4. Khan, Adil Mehmood and Lee, Young-Koo and Lee, Sungyoung Y and Kim, Tae-Seong, A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer, Information Technology in Biomedicine, IEEE Transactions on,2010;14:5-1166.
5. Lester, Jonathan and Choudhury, Tanzeem and Borriello, Gaetano, A practical approach to recognizing physical activities, Pervasive Computing, 2006; Springer
6. Kwapisz, Jennifer R and Weiss, Gary M and Moore, Samuel A, Activity recognition using cell phone accelerometers, ACM SigKDD Explorations Newsletter, 2011;12:2-74.
7. Ravi, Nishkam and Dandekar, Nikhil and Mysore, Preetham and Littman, Michael L, Activity recognition from accelerometer data, AAAI, 2005;5-1541.
8. Mannini, Andrea and Sabatini, Angelo Maria, Machine learning methods for classifying human physical activity from on-body accelerometers, Sensors, 2010;10:2-1154.
9. Ruta, Dymitr and Gabrys, Bogdan, Classifier selection for majority voting, Informtion fusion, 2005;6:1-63.
10. Gyllensten, Illapha Cuba and Bonomi, Alberto G, Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life, Biomedicl Engineering, IEEE Trnsaction on, 2011;58:9-2656.
11. Alelyani, Salem and Tang, Jiliang and Liu, Huan, Feature selection for clustering: A review, Data Clustering: Algorithms and Applications, 2013;29.
12. Fujiki, Yuichi, iPhone as a physical activity measurement platform, CHI'10 Extended Abstracts on Human Factors in Computing Systems, 2010;4315.
13. Krishnan, Narayanan C and Colbry, Dirk and Juillard, Colin and Panchanathan, Sethuraman, Real time human activity recognition using tri-axial accelerometers, Sensors, signals and information processing workshop, 2008.
14. Foerster, Friedrich and Fahrenberg, Jochen, Motion pattern and posture: correctly assessed by calibrated accelerometers, Behavior research methods, instruments, & computers, 2000;32:3-450.
15. Westerterp, Klaas R, Assessment of physical activity: a critical appraisal, European journal of applied physiology, 2009;105:6-823.
16. Poppe, Ronald, A survey on vision-based human action recognition, Image and vision computing, 2010;28:6-976.
17. Bodor, Robert and Jackson, Bennett and Papanikolopoulos, Nikolaos, Vision-based human tracking and activity recognition, Proc. of the 11th Mediterranean Conf. on Control and Automation, 2003.
18. Kwapisz, Jennifer R and Weiss, Gary M and Moore, Samuel A, Cell phone-based biometric identification, Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, 2010.