An Efficient Feature Selection Method for Activity Classification

Shumei Zhang[1]
1. Department of Computer Science
Shijiazhuang University, China
E-mail: zhang-s2@email.ulster.ac.uk

Paul McCullagh[2],Vic Callaghan[3]
2. University of Ulster, Northern Ireland, UK
3. University of Essex, UK
E-mail: pj.mccullagh@ulster.ac.uk, vic@essex.ac.uk

***Abstract*—Feature selection is a key step for activity classification applications. Feature selection selects the most relevant features and considers how to use each of the selected features in the most suitable format. This paper proposes an efficient feature selection method that organizes multiple subsets of features in a multilayer, rather than utilizing all selected features together as one large feature set. The proposed method was evaluated by 13 subjects (aged from 23 to 50) in a lab environment. The experimental results illustrate that the large number of features (3 vs. 7 features) are not associated with high classification accuracy using a single Support Vector Machine (SVM) model (61.3% vs. 44.7%). However, the accuracy was improved significantly (83.1% vs. 44.7%), when the selected 7 features were organized as 3 subsets and used to classify 10 postures (9 motionless with 1 motion) in 3 layers via a hierarchical algorithm, which combined a rule-based algorithm with 3 independent SVM models.**

Keywords-feature selection; signal analysis; hierarchical algorithm; activity classification; smart phone

I. INTRODUCTION

Physical activity has the potential to reduce the risk of many chronic diseases [1]. Daily activity monitoring techniques that can be used to encourage people to adopt a healthier lifestyle and engage in further activities if required, have the potential to offer significant savings in future healthcare costs especially for the elderly and those who are suffering from a form of chronic disease [2].

Feature selection is a key step for activity classification applications. Feature selection aims to select the minimum subset of features for improving classification accuracy and resulting class distribution. The values for the selected features should be as close as possible to the original class distribution, given all feature values [3]. The features are selected for each activity by identifying one feature having the best performance that distinguishes the required activity from other activities. Features are derived from the raw sensing data directly or by performing calculations on the raw data. Features selection adheres to the following principles: a) Discrimination: features should have significantly different values for different classes; b) Reliability: features should have similar values for the same class; c) Independence: features should not be influenced strongly to each other; d) Optimality: features should not be redundant; e) Small: the number of features should be as small as possible to reduce the complexity of the classifier [4].

This study aims to design an efficient strategy for optimal feature selection from the accelerometers embedded in a smart phone, and to improve the real-time activity classification accuracy.

II. RELATED WORK

Bao & Intille [5] extracted four types of Fast Fourier Transform -based features from acceleration data, which were sensed using five biaxial body-worn accelerometers. Twenty different daily activities were classified based on the four selected features, which included means, energy, frequency-domain entropy, and correlation of acceleration data. Then several classifiers were tested using these features. Their experimental results demonstrated that the best result was obtained using the decision tree classifier, and good performance was indeed achieved using only 2 of the 5 accelerometers (thigh and wrist).

Ermes et al. [6] selected 7 signal features for 10 different activities for classification based on two accelerometers. The 7 features included: Peak frequency, Range, Mean value, Sum, and Spectral entropy of the acceleration measured from the hip; Peak frequency of the horizontal acceleration measured from the wrist; Speed calculated from the GPS information. The performance of each feature was assessed using the receiver operator characteristic (ROC) curve. The experimental results based on the selected 7 features illustrated that a 90% recognition rate was achieved for supervised data. The accuracy was reduced by 17% when supervised data were used for training and only unsupervised data for evaluation.

Mäntyjärvi et al. [7] employed Principle Component Analysis (PCA) and Independent Component Analysis (ICA) algorithms to extract a $6 \times 4 = 24$ dimensional feature vector from a pair of 3D sensors, attached to the left and right hips. They classified 4 different activities using a neural networks classifier. The activity classifications were examined based on three data sets respectively; original sensing data, PCA and ICA feature vectors. Their experimental results indicated that the classification accuracy was similar using ICA or PCA feature vectors, which achieved 83-90%. Nevertheless, the result was 61-84% using the raw data set.

Maguire & Frisby [8] identified eight activities using two classifiers (k-NN and J48) based on combined accelerometer and heart rate data.. Thirteen features were initially selected (mean, standard deviation, energy and correlation from each of the 3-axes acceleration, and mean heart rate). The effect of removing features on classification accuracy using the best first search method was investigated. Finally, the thirteen features were reduced to seven (meanHeartRate, meanZ, stdX, stdY, stdZ, energyY and energyZ) with minimal impact on classification accuracy. Ravi et al [9] extracted four features from the raw accelerometer data using a window size of 256 with 128 samples overlapping between consecutive windows. The features selected were: Mean, Standard deviation, Energy, and Correlation.

978-1-4799-2947-4/14 $31.00 © 2014 IEEE
DOI 10.1109/IE.2014.10

Usually, feature extraction/selection techniques have the potential to improve the classification accuracy, but at a cost in computational time [10]. For real-time applications, we need to consider a method that could provide the best possible classification accuracy, and also could have high computational efficiency.

III. METHODS

A. Data Sensing

An HTC smart phone was used for data sensing and processing in this study. The phone comprised embedded BMA150 3D accelerometer, orientation sensor, 3D Magnetic sensor, GPS and Wi-Fi. The phone's processor operates at 600MHz, the memory capacity is 512MB with an additional 2GB memory card and the operating system is Android version 2.3.3.

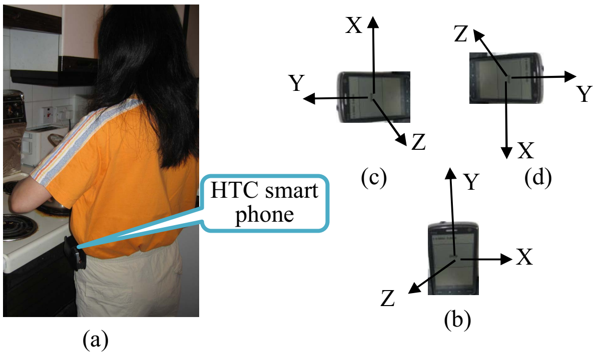

Figure 1. System configuration; (a) the phone is belt-worn horizontally on the left side of the waist; (b) Y-axis up if the phone is vertical; (c) & (d) X-axis up/down if the phone is horizontal and facing backward/frontward.

The phone is belt-worn on the left side of the waist in a horizontal orientation as shown in Fig.1 (a). The sensor coordinate system is defined relative to the phone's screen as shown in Fig.1 (b), (c) and (d).

The data set of 3D acceleration (*t, Ax, Ay, Az*) was obtained by using the accelerometer sensor embedded in the smart phone. Subsequently, features were extracted from the raw acceleration signals, and used for the evaluation of the activity classification accuracy based on different features using machine learning algorithms. The recorded data were saved in the phone in text format.

B. Sampling frequency setting

In theory, by using a low sampling rate it is possible to miss some peak values for motion activities (such as walking). Although the missed values can influence the detailed analysis for motion activities, such as walking speed, it should not influence on the accuracy motion and motionless postures classification.

In order to verify the relationship between sampling frequency *f* and classification accuracy, two frequencies 1Hz vs. 10Hz were compared in two aspects: data processing time and classification accuracy. Two smart phones were set in different sampling frequency, one was set f = 1Hz, and another was set f = 10 Hz. One subject wears these two phones on the left waist at the same time and did a series of activities as:

Start phone1-start phone2-walk- jump- walk- standing- walk slowly- sitting- walk fast- standing- walk (short)- jump- walk- sitting- stop phone1-stop phone2.

The experimental results for f = 1Hz vs. f = 10Hz are shown in Fig. 2. This illustrates that the sampling frequency f= 10Hz results in a larger number of samples (3222 vs 339 for 5 minutes data recording), but does not influnce the activity classification accuracy compared to the results in f=1Hz. Although some higher peak values were missed using f=1Hz sampling, the postures such as 'jump' can still be distinguished from 'walk' postures. The lower sampling rates result in a lower data load and higher efficiency of data processing.

Therefore the sampling frequency was set at 1Hz in this study to reduce the data load and improve the efficiency of data processcing for real-time applications.

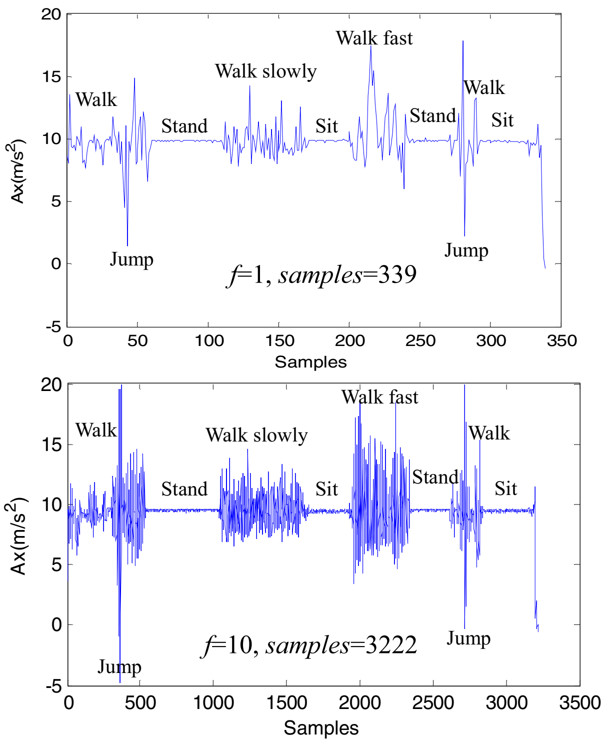

Figure 2. Comparison of acceleration Ax signals and activity classification results between f=1 and f=10. The sample number was 339 vs 3222 for 5 minutes data recording, however the posture classification results shown in both figures were same that classified and marked using SVM model automaticlly in Matlab environment.

C. Features extraction

A computationally expensive feature extraction method is not suitable for resource constrained and battery powered devices [9]. Mobile phones are weaker in terms of processing power. They also lack the constant supply of power and must rely on their limited battery resource [10]. Hence the feature extraction techniques such as FFT [5], ICA, or PCA [7] are computation-intensive and are not suitable for smart phone based real-time applications.

Considering the efficiency of data processing for real-time system, the following features were extracted from the 3-axes accelerometer sensor in this study.

- The raw 3-axes acceleration with time stamp organized as (*t, Ax, Ay, Az*);

- The three dimensional acceleration A and the acceleration change ΔA. Where A and ΔA are calculated using (1) and (2) respectively;
- The tilt angles of the accelerometer around 3-axes (ρ, φ, θ). Here *Pitch* (ρ) is defined as the angle of the X-axis relative to the ground; *Roll* (φ) is defined as the angle of the Y-axis relative to the ground; *Theta* (θ) is the angle of the Z-axis relative to gravity.

$$\mathrm{A} = \sqrt{A_x^2 + A_y^2 + A_z^2} \qquad (1)$$

$$\Delta\mathrm{A} = |\mathrm{A}(t_{i+1}) - \mathrm{A}(t_i)| \qquad (2)$$

Tilt angle uses measurements of gravity and its trigonometric projection on the axes of a 3-axial accelerometer, to determine tilt angle in all three spatial dimensions (*X*, *Y* and *Z*). Three angles (ρ, φ, θ) can be calculated using (*Ax, Ay, Az*) as shown in (3), (4) and (5).

$$\rho = {}^{180}/_{Pi} \times \arctan\left(\frac{A_x}{\sqrt{A_y^2 + A_z^2}}\right) \qquad (3)$$

$$\varphi = {}^{180}/_{Pi} \times \arctan\left(\frac{A_y}{\sqrt{A_x^2 + A_z^2}}\right) \qquad (4)$$

$$\theta = {}^{180}/_{Pi} \times \arctan\left(\frac{A_z}{\sqrt{A_y^2 + A_x^2}}\right) \qquad (5)$$

$$where\ Pi(radians) = 180^{\circ}$$

Fig.3 (a) shows the original three dimensional acceleration coordinate; Fig.3 (b) shows the accelerometer coordinate system is that the X-axis is vertical, the Y-axis is horizontal and the Z-axis is orthogonal to the screen when the phone is belt-worn on the waist; Fig. 3 (c) shows the three tilt angles.

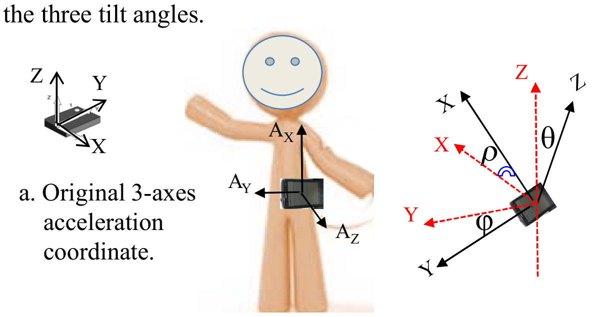

Figure 3. Three tilt angles for measuring tilt postures.

Acceleration is a physical characteristic of a subject in motion. An accelerometer is a device that can measure the static acceleration due to gravity, and dynamic acceleration resulting from motion, shock, or vibration [11]. An accelerometer will measure a value of ±1g (unit of gravity acceleration, which is 9.81m/s^2) in the upward or downward direction if it remains stationary relative to the earth's surface.

D. Existed features selection approaches

A redundant feature does not contribute anything new to the classifier, and an irrelevant feature does not improve the classification result. Including weak or irrelevant features not only slows computation, but can also degrade classification performance [12]. Feature selection methods try to pick a small subset of the most relevant features that can improve the performance of the classifiers and increase computational efficiency.

General feature selection approaches are the "top-down" and "bottom-up" [13]. The top-down method is initialized with the whole feature set, and features are removed based on a certain criteria. Finally the subset of remaining features is selected as the best subset [14]. The bottom-up method is initialized with the empty set, and features are added until a predefined certain criteria have been met [15]. The disadvantage of these methods is the nesting effect, in which features removed are no longer considered or features added are never removed.

Different heuristic approaches are used to avoid the nesting effect. Most heuristic feature selection approaches can be categorized into two types: filter and wrapper methods [16]. Filter-based methods select the subset of features by employing statistical evaluation algorithms such as autocorrelation analysis, spectral analysis, or stepwise regression. Therefore, filters can limit the search space of possible meta-parameters, while it is independent of a particular classification algorithm. Wrapper methods select the subset of features according to the resulting predicting accuracy by using an underlying classification model. Wrappers are often employing a grid-search or an exhaustive evaluation of meta-parameters to identify best meta-parameters [17]. Obviously, the best feature subset is usually selected through the wrapper methods if a classifier is given. However, wrapper methods are often criticized for their computational inefficiency caused by the joint processing of learning and feature selection tasks [18].

In this paper, a signal analysis method is used to select multiple subsets of features and apply those respectively using hierarchical classifiers to improve the classification accuracy and computational efficiency.

E. Signal analysis strategy for feature selection

There are 9 features extracted in total, which include the raw data (t, *Ax, Ay, Az*) sensed using a smart phone embedded accelerometer with features (Δt, ΔA) & (ρ, φ, θ) calculated based on the raw data. The problem is how to identify each of the extracted features having the best performance to distinguish one type of activity from other activities.

Signal analysis aims to analyze its strength and weakness for each of the features for the discriminating classes by analyzing whether the feature has significantly different values for different classes, and whether it also has similar values for the same class.

The signals for each of the acceleration features (*Ax, Ay, Az, A,* ΔA) are compared based on different activity postures as shown in Fig.4.

Fig.4 illustrated that the three dimensional acceleration (*Ax, Ay, Az*) can be used to classify most of activity postures such as sit, stand, lying and walk, however some of these postures have similar values, for instance, *sit-N, sta-U*, with *sta-F* and *sit-L* with *sit-B* are difficult to classify properly using only three (*Ax, Ay, Az*) features. Additionally, most of *walk* instances have similar acceleration values with *stand* instances, which also is difficult to distinguish both correctly if the classification is only based on the feature subset (*Ax, Ay, Az*). Obviously, the feature ΔA or feature A can be used to separate motion and motionless postures clearly.

In this study, the motion and motionless postures are defined as below:

- *If the change of acceleration (ΔA) is continued more than an empirical value 0.8 m/s^2 for more than 2 seconds (Δt), then this activity can be defined as motion.*
- *If the change of acceleration (ΔA) is continued less than an empirical value 0.4 m/s^2 for more than 5 seconds (Δt), then this activity can be defined as motionless.*

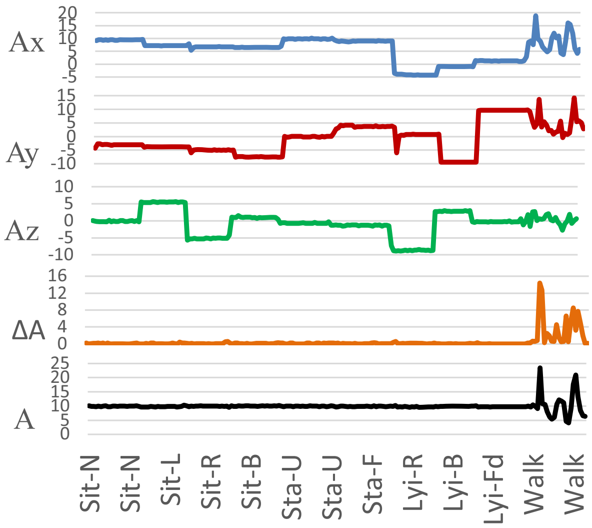

Figure 4. Signals for each of the acceleration features based on selected activity postures

The signals for each of the tilt angle features (ρ, φ, θ) are shown in Fig.5. Fig.5 demonstrated that the subset of features (ρ, φ, θ) can be used to further classify the different sitting postures such as *sit-N, sit-L, sit-R* and *sit-B*, or different standing postures.

According to the definition of the (ρ, φ, θ) above, the three angles will vary according to the specific body postures. For example, when the body posture is stand upright (*sta-U*), the X-axis is vertical, then $\rho \approx 90°$ and $\varphi \approx \theta \approx 0°$; otherwise, when the body posture is sit normal (sit-N), then it is definitely $|\rho| > |\varphi| > |\theta|$, in theory. Of course, the theoretic values will have a small fluctuation in the real life.

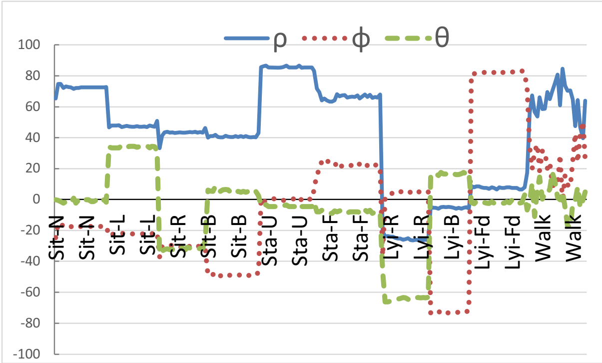

Figure 5. The three tilt angles' signal derived from 10 daily activity postures. Sit-N is sit normal or upright; Sit-L, Sit-R and Sit-B are sit tilted left, right and back respectively; Sta-U and Sta-F are stant upright and stant forward respectively; Lyi-R, Lyi-B and Lyi-Fd are lying right, back and face down respectively; walk is walking.

F. *Using multiple subsets of features hierarchically*

According to the signal analysis, three feature subsets are selected from the extracted 9 feature set (t, *Ax, Ay, Az, A*, ΔA, ρ, φ, θ) named as *F1*= (Δt, ΔA), *F2*=(*Ax, Ay, Az*), and *F3*=(ρ, φ, θ).

Then the three subsets of features are applied hierarchically using a rule-based classifier with 3 different SVM models (SVM-ml, SVM-m, and SVM-sit). The details of the procedure are described as shown in Fig.6.

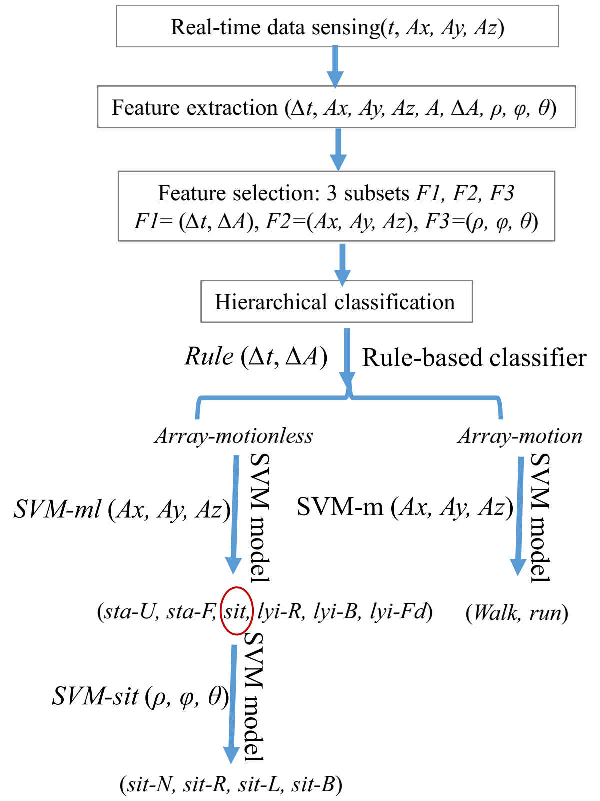

Figure 6. Selected 3 subsets of features are applied hierarchically based on rule-based classifier with three different SVM models.

(1) Based on signal analysis from Fig.4, the subset (Δt, ΔA) of features is selected to discriminate motion and motionless postures firstly using a rule-based classifier.

Obviously, the *F1* (Δt, ΔA) is not able to further classify the different motionless postures such as between stand, sit and lying, since all motionless postures have similar ΔA values as Fig.4 shown.

(2) Based on the signal analysis from Fig.4, the feature subset (*Ax, Ay, Az*) is selected to classify the different motionless postures and different motion postures respectively using two SVM models SVM-ml and SVM-m. Nevertheless, it is limited to distinguish different sitting postures such as *sit-N, sit-R, sit-L,* and *sit-B*.

(3) Based on the signal analysis from Fig.5, the subset (ρ, φ, θ) of features is selected to further classify different sitting postures using SVM-sit model, if the postures is classified as sit using SVM-ml.

The four classifiers *Rule* (Δt, ΔA), *SVM-ml* (*Ax, Ay, Az*), *SVM-m* (*Ax, Ay, Az*), and *SVM-sit* (ρ, φ, θ) are trained based on relevant training data set respectively. Details about the SVM model training methods were described in our previous work [19].

IV. Experiments

The experiments were performed by 13 subjects (aged from 23 to 50) in a laboratory based setting where a video was fixed in the ceiling. The experimental results were validated by synchronized video recordings. Data collected from subject1 were used as the training set *trainD*, and the data collected from the other 12 subjects were used as test set *testD*. In order to compare the testing accuracy between different subset of features and different classification algorithms, the data sets were organized as several pair of training and test sets based on different feature combination, for example, three dimension acceleration (*Ax, Ay, Az*), three dimension tilt angle (ρ, φ, θ), three dimensional acceleration change (Δt, ΔA), or all 7 features (*Ax, Ay, Az,* ΔA, ρ, φ, θ) together. In addition, both train and test sets were also organized in different groups based on different postures combination such as all motion and motionless postures together, only motion and motionless postures respectively. In this paper, each of the 13 subjects did 10 postures continuously that labled as (*sit-N, sit-R, sit-L, sit-B, sta-U, sta-F, lyi-R, lyi-B, lyi-Fd, walk*). Here *sit-N, sit-R, sit-L, sit-B* signify sit normal/upright, sit leaning right, left and back respectively; *sta-U* and *sta-F* mean stand upright and tilt forward; *lyi-R, lyi-B,* and *lyi-Fd* imply lying right, on back and face down.

We have developed a real-time activity classification system using a hierarchical algorithm to separate data set and classify different data set automatically within a smart phone [19][20]. All classifications were performed in the RapidMiner environment.

A. Small feature subset vs. all features together (3 vs.7)

Optimal feature subsets should not be redundant, and size is preferred as small as possible to reduce the complexity of the classifier. Therefore, the feature selection method not only selects the relevant features, but also considers how to use each of the selected features in the most suitable format.

Activity classification accuracy is compared between two feature subsets using a SVM classifier. One of the feature subset is selected in small size (3 features) based on the signal of features analysis; and another is simply putting all features together (7 features).

All data collected from 12 test subjects that include motion and motionless postures labeled in total 10 classes were organized together, and classified based on 3 features (*Ax, Ay, Az*) and 7 features (*Ax, Ay, Az,* ΔA, ρ, φ, θ) respectively, using a trained single SVM model. The classification results are show in Table 1 and Table2.

Results indicate that the larger number of features are not associated with higher classification accuracy. The average accuracy is 61.3% for the subset of 3 features, and 44.7% for the 7 features.

When the classification is performed on the subset of 3 features (*Ax, Ay, Az*), many motionless samples, 306 out of 477 sta-F instances, and 154 out of 482 lyi-Fd instances were classified as walk; 119 out of 508 walk instances were classified as sta-F as shown in Table 1. However, when the classification is performed on the subset of 7 features, 3 of the 10 classes accuracy were improved such as walk (89.4% vs. 64.2%), lyi-Fd (91.7% vs. 66%) and lyi-R (93.7% vs. 77.3%), however 7 of the 10 classes accuracy were decreased, especially sit-R (12.2% vs. 94.8%) and sit-L (0.3% vs. 39.6%) as shown in Table 2.

TABLE I. CLASSIFICAION RESULT USING A SINGLE SVM MODEL BASED ON 3 FEATURES (A_X, A_Y, A_Z)

True:	sit-N	sit-R	sta-U	sit-B	sit-L	sta-F	lyi-B	lyi-R	lyi-Fd	walk
sit-N	230	0	10	26	68	0	0	0	0	14
sit-R	167	312	0	144	71	0	75	0	0	0
sta-U	29	3	364	0	0	16	0	0	0	37
sit-B	0	14	0	182	87	0	35	0	0	0
sit-L	0	0	2	25	148	0	0	0	0	0
sta-F	2	0	74	1	0	155	0	45	0	119
lyi-B	0	0	0	0	0	0	378	56	0	1
lyi-R	0	0	0	0	0	0	0	343	10	11
lyi-Fd	0	0	0	0	0	0	0	0	318	0
walk	0	0	168	0	0	306	0	0	154	326
class Acc.	53.7%	94.8%	58.9%	48.1%	39.6%	32.5%	77%	77.3%	66.0%	64.2%
Average accuracy		61.3%								

TABLE II. CLASSIFICAION RESULT USING A SINGLE SVM MODEL BASED ON 7 FEATURES (A_X, A_Y, A_Z, ΔA, P, Φ, Θ)

True:	sit-N	sit-R	sta-U	sit-B	sit-L	sta-F	lyi-B	lyi-R	lyi-Fd	walk
sit-N	110	1	43	0	12	0	0	0	0	19
sit-R	99	40	0	142	113	0	65	0	0	0
sta-U	2	1	81	0	0	0	0	0	0	6
sit-B	6	1	0	104	77	0	0	0	0	1
sit-L	0	0	0	0	1	0	32	0	0	0
sta-F	211	285	127	131	38	140	0	0	0	21
lyi-B	0	0	0	0	0	0	312	0	0	0
lyi-R	0	0	0	0	0	0	78	416	0	7
lyi-Fd	0	0	0	0	0	0	1	28	442	0
walk	0	1	367	1	133	337	0	0	40	454
class Acc.	25.7%	12.2%	13.1%	27.5%	0.3%	29.4%	63.9%	93.7%	91.7%	89.4%
Average accuracy		44.7%								

The single SVM model achieves average classification accuracy between motion and motionless samples based on both 3 and 7 selected features (61.3% vs. 44.7% respectively).

B. The hierarchical approach vs. single classifier

The hierarchical classification approach, described in Fig.6 is evaluated in this section.

Foremost, the subset of features (Δt, ΔA) was used classify the sensed original data into two arrays (*motionD and mlessD*) by a rule-based algorithm *Rule* (Δt, ΔA). Then the subset of features (*Ax, Ay, Az*) was used to further classify the motion array *motionD* and the motionless array *mlessD* respectively by two SVM models *SVM-m* and *SVM-ml*. For example, the average classification accuracy for motionless postures is 78.1% as shown in Table 3.

Table 3 demonstrated that the classification accuracy for most of the 9 motionless classes were improved compared to the approach of using all selected 7 features together (in Table 2), such as *sit-N* (84.6% vs. 25.7%), *sta-U* (80.5% vs. 13.1%), *sta-F* (99.6% vs. 29.4%), *lyi-B* (100% vs. 63.9%). The average accuracy was improved also (78.1% vs. 44.7%). However, there were two classes *sit-B* and *sit-L* still got a lower accuracy (46.5% and 35%), since the classifier confused among the 4 sit postures (*sit-N, sit-R, sit-B* and *sit-L*).

In order to solve the problem of confusion among different sit postures, the 4 sit classes were relabeled as one class '*Sit*' in the motionless array *mlessD*. Thus, the 9 motionless postures were classified in two-level again. Firstly, the integrated 6 motionless postures was classified by re-trained *SVM-ml* model based on the subset of features (*Ax, Ay, Az*), and obtained a higher average classification accuracy (93.8%) as shown in Table 4; then the 4 types of *Sit* classes were further classified as 4 different sit postures by the *SVM-sit* model based on the subset of features (ρ, φ, θ), and the classification accuracy for each of the sit postures was improved (except the sit-R), compared to classify all 9 motionless postures together, as shown in Table 5 (72% vs. 57.9% in average).

Finally, the classification accuracy for each of the 9 motionless postures, as well as the average accuracy was calculated by combining Table 4 and Table 5, the result was shown in Table 6. Obviously, the hierarchical approach can improve the accuracy of activity classification significantly compared to using all selected features together via a single classifier (83.1% vs. 44.7% in average).

TABLE III. CLASSIFICATION RESULT FOR ALL MOTIONLESS CLASSES, BASED ON FEATURE ((A_X, A_Y, A_Z) (USING A SINGLE SVM)

True:	sit-N	sit-R	sta-U	sit-B	sit-L	sta-F	lyi-B	lyi-R	lyi-Fd
sit-N	252	1	3	1	85	0	0	0	0
sit-R	4	146	0	74	1	0	0	0	0
sta-U	41	0	354	0	0	1	0	0	0
sit-B	0	41	0	101	60	0	0	0	0
sit-L	0	0	0	41	78	0	0	0	0
sta-F	1	0	83	0	0	256	0	0	0
lyi-B	0	1	0	0	0	0	258	20	0
lyi-R	0	0	0	0	0	0	0	191	0
lyi-Fd	0	0	0	0	0	0	0	28	225
Acc.	84.6%	77.3%	80.5%	46.5%	35%	99.6%	100%	80%	100%
Average accuracy		78.1%							

TABLE IV. CLASSIFICATION RESULT FOR MOTIONLESS CLASSES WITH INTEGRATED SIT POSTURES BASED ON FEATURE (A_X, A_Y, A_Z)

True:	Sit	lyi-Fd	lyi-B	sta-F	lyi-R	sta-U
Sit	1444	0	1	0	0	5
lyi-Fd	0	488	0	0	28	0
lyi-B	0	0	492	0	20	0
sta-F	2	0	0	439	0	85
lyi-R	0	0	0	0	400	0
sta-U	45	0	0	41	0	534
Acc.	96.8%	100.0%	99.8%	91.5%	89.3%	85.6%
Average accuracy		93.8%				

TABLE V. FURTHER CLASSIFICATION RESULT FOR ONLY SIT POSTURES BASED ON FEATURE (P, Φ, Θ)

True:	sit-N	sit-B	sit-R	sit-L	Acc. Based on (A_x, A_y, A_z)
sit-N	348	2	58	47	84.6%
sit-B	0	229	27	80	46.5%
sit-R	2	68	203	0	77.3%
sit-L	45	31	6	201	35%
Acc.	88.1%	69.4%	69.0%	61.3%	Above copied from Tab.3
Average accuracy		72.0%			57.9%

TABLE VI. THE FINAL CLASSIFICATION ACCURACY FOR ALL 9 MOTIONLESS CLASSES, CALCULATED BY COMBINING TABLE 4 AND 5 (USING TWO HIERARCHICAL SVM-ML AND SVM-SIT).

classes	sit-N	sit-B	sit-R	sit-L	sta-U	sta-F	lyi-B	lyi-R	lyi-Fd
Acc.	88.1%	69.4%	69%	61.3%	80.5%	99.6%	100%	80.0%	100%
Average accuracy		83.1%							

V. CONCLUSION AND FUTURE WORK

This paper proposed an efficient feature selection method that combines a simple signal analyzing approach with the concept of organizing multiple small subsets of features in the multilayer, rather than grouping all selected features together as a larger feature set.

Feature selection aims to select the most relevant input variables within a dataset, which should have significantly different values for different classes, and have similar values for the same class. An optimal feature subset with non-redundant features can improve the performance of the classifiers and increase the computational efficiency. Each of the features has its strength and weakness in discriminating some of classes, hence the feature selection method not only selects the relevant features, but also considers how to use each of the selected features in the most suitable format.

The familiar existing heuristic approaches for feature selection are categorized into two types: filter and wrapper. However, no single methodology has been proven to perform well consistently across varying data conditions, given their individual shortcomings. For example, wrapper methods are often criticized for their computational inefficiency.

The proposed signal analysis approach can visually identify each of the extracted features in terms of its discriminating power (i.e. which type of classes to distinguish definitely), without increased computational processing costs. So it is efficient. The concept of using multiple small subsets of features hierarchically can improve the classification accuracy. The proposed feature selection method was evaluated based on data collected from 13 subjects (one as training, the other 12 as test subjects).

Two experiments were designed to compare the classification accuracy between different subset of features (3 vs. 7 features) using a SVM model, and between a single SVM and a hierarchical classifier that integrated a rule-based algorithm with 3 independent SVM models. The experimental results illustrated that the large number of features are not associated with high classification accuracy via a SVM model (61.3% vs. 44.7%). However, the accuracy was improved significantly (83.1% vs. 44.7%), when the selected 7 features were organized as 3 subsets and used to classify 10 postures (9 motionless with 1 motion) in 3 layers via a hierarchical algorithm.

The future work will evaluate the algorithms using more daily activity postures, especially more motion activities such as walk, run, up stair, down stair, and driving. It is also important to evaluate the algorithms for other subjects groups, such as the elderly.

ACKNOWLEDGMENT

This work was supported by Natural science foundation of Hebei province China, grants No. F2013106121 and the High-level talents in Hebei province funded project for overseas student activities of science and technology, grants No. C2013003036. The authors also acknowledge all members of the Smart Environments Research Group of Ulster University in UK for their help with collecting the experimental data.

REFERENCES

[1] Blair, S.N. 2009, "Physical inactivity: the biggest public health problem of the 21st century", *British journal of sports medicine,* vol. 43, no. 1, pp. 1.

[2] Bates, D.W. & Bitton, A. 2010, "The future of health information technology in the patient-centered medical home", *Health affairs,* vol. 29, no. 4, pp. 614-621.

[3] Dash, M., & Liu, H. (1997). Feature selection for classification. Intelligent data analysis, 1(3), 131-156.

[4] Kim, J. K., & Park, H. (1999). Statistical textural features for detection of microcalcifications in digitized mammograms. Medical Imaging, IEEE Transactions on, 18(3), 231-238.

[5] Bao, L., and Intille, S. S. 2004, “Activity recognition from userannotated acceleration data”, In Proceceedings of the 2nd International Conference on Pervasive Computing, 1–17.

[6] Ermes, M., Parkka, J., Mantyjarvi, J. & Korhonen, I. 2008, "Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions", *IEEE Transactions on Information Technology in Biomedicine,* vol. 12, no. 1, pp. 20-26.

[7] Mantyjarvi, J., Himberg, J. & Seppanen, T. 2002, "Recognizing human motion with multiple acceleration sensors", *IEEE International Conference on Systems, Man, and Cybernetics,* pp. 747-752.

[8] Maguire, D., & Frisby, R. (2009, October). Comparison of feature classification algorithm for activity recognition based on accelerometer and heart rate data. In *9th. IT & T Conference* (p. 11).

[9] Ravi, N., Dandekar, N., Mysore, P. & Littman, M.L. 2005, "Activity recognition from accelerometer data", Proceedings of the National Conference on Artificial IntelligenceMenlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; pp. 1541.

[10] Győrbíró, N., Fábián, Á., & Hományi, G. (2009). An activity recognition system for mobile phones. Mobile Networks and Applications, 14(1), 82-91.

[11] Duda R O, Hart P E, Stork D G. Pattern classification[M]. John Wiley & Sons, 2012.

[12] Duda R O, Hart P E, Stork D G. Pattern classification and scene analysis. Part 1, Pattern classification[J]. 2000.

[13] Way, T. W., Sahiner, B., Hadjiiski, L. M., & Chan, H. P. (2010). Effect of finite sample size on feature selection and classification: A simulation study. Medical physics, 37(2), 907-920.

[14] Marill, T., & Green, D. M. (1963). On the effectiveness of receptors in recognition systems. IEEE Transactions on Information Theory, 9(1), 11-17.

[15] Whitney, A. W. (1971). A direct method of nonparametric measurement selection. IEEE Transactions on Computers, 100(9), 1100-1103.

[16] Unler, A., & Murat, A. (2010). A discrete particle swarm optimization method for feature selection in binary classification problems. European Journal of Operational Research, 206(3), 528-539.

[17] Crone, S. F., & Kourentzes, N. (2010). Feature selection for time series prediction–A combined filter and wrapper approach for neural networks. Neurocomputing, 73(10), 1923-1936.

[18] Reunanen, J. (2003). Overfitting in making comparisons between variable selection methods. The Journal of Machine Learning Research, 3, 1371-1382.

[19] Zhang, S., McCullagh, P., Nugent, C., Zheng, H. & Baumgarten, M. 2011, "Optimal model selection for posture recognition in home-based healthcare", *International Journal of Machine Learning and Cybernetics,* vol. 2, no. 1, pp. 1-14

[20] Zhang, S., McCullagh, P., Zhang, J., Yu, T., 2014, “A smartphone based real-time daily activity monitoring system”, Journal of Cluster Computing, online first, DOI: 10.1007/s10586-013-0335-y.