Adaptive Energy-Saving Strategy for Activity Recognition on Mobile Phone

Vo Quang Viet
ECE, Chonnam National University
Gwangju, South Korea
Email: vietquangvo@gmail.com

Hoang Minh Thang
ECE, Chonnam National University
Gwangju, South Korea
Email: hmthang2812@gmail.com

Deokjai Choi
ECE, Chonnam National University
Gwangju, South Korea
Email: dchoi@jnu.ac.kr

Abstract— **Most existing mobile devices nowadays are powered by a limited energy resource. With the tendency using machine learning on mobile devices for activity recognition (AR), recent achievements still remain restrictions including low accuracy and lacking of evidences about power consumption of feature extraction and classification. Moreover, keeping constantly a high sampling frequency was the most power consuming factor. In this paper, we contribute a novel method for extracting features in time domain and frequency domain. These features are then classified by Support Vector Machine (SVM). Prototypes of the proposed methods are then implemented on a cell phone to measure power consumptions. To reduce the energy overhead of continuous activity recognizing, we propose an adaptive energy-saving strategy by selecting an appropriate combination of flexible frequency and classification feature for each individual. The self-construct data and SCUTT-NAA dataset are used in our experiment. We achieved an overall 28 percent of energy saving in activity recognition on mobile phone**.

Keywords- Mobile Accelerometer; Activity Recognition; SVM Classifier; Power Consumption; Adaptation Strategy

I. INTRODUCTION

Accelerometer-based AR is not a new topic. In healthcare applications, it has been used to detect a fall and the movements of user after falling, predict user's energy consumption based monitoring activity of daily living (ADL), and generate daily, weekly, and monthly activity reports in order to promote health and fitness. In context-aware pervasive computing systems, mobile accelerometer has been gained significant achievements. Activity information can be used to adapt behavior of using mobile phone automatically. It can include sending calls directly to voicemail if a user is bicycling or jogging, turning on music when jogging is taken place, etc. In this study, we only focus on mobile approach as a standalone device without using extra resources since sending the data to central server from mobile device can generates privacy issues [1].

Some mobile achievements have been attempting to recognize ADLs as in [2-3]. Normally, AR is passed in three steps. First, data windows from segmentation of accelerometer signal are taken. Second, some features that describe the clearest properties of studying activity are extracted. These preprocess steps are the most important parts of AR system since the last step is classification that can be studied by any existing machine learning algorithm.

From the best of our knowledge, previous achievements in this area also remain restrictions including:
- Accuracies of achievements were not stable especially in predicting of multiple-subject type [2, 4].
- Power consumption was not reported.
- Most of achievements were collected from a fixed sampling frequency of mobile accelerometer [3, 4].

In this paper we study on walking, bicycling, jogging, up-stairs, and down-stairs activities. Mobile phone is placed in the front pant pocket because of the popularity of attaching phone at this position [5]. Applying AR on mobile phone is unlike applying on central server. It not only requires the effective accuracy but also guarantees the lowest power consumption. Therefore, our work is carried out as following in order to guarantee these standards:

- Propose a novel method for feature extraction. It includes effectively using Y, Z axes from mobile tri-axial accelerometer, and segmenting on main Y-axis based peak detection algorithm. Features in each window are extracted in time domain and frequency domain. These features are then classified using SVM classifier respectively. In order to guarantee an effective accuracy, these methods are applied on SCUT-NAA standard dataset [6] and our own data collected from mobile accelerometer. Output results are compared with other previous achievements in individual and multiple subject types.
- Our novel feature extraction is then applied to distinct sampling frequencies (SF) of mobile accelerometer to compare the classification accuracy and energy consumption. After that, an adaptive strategy is proposed by selecting an appropriate combination of flexible SF and classification feature (CF) for each activity. This work is then experimented to guarantee an effective method for energy saving in AR using mobile accelerometer.

II. A NOVEL METHOD FOR FEATURE EXTRACTION

We paid particular attention to the position of accelerometer. The mobile phone was vertically fixed at the pocket location as shown in figure 1. From three axes of accelerometer, the X-axis captures horizontal movement of the user's leg. The Y-axis captures the upward and

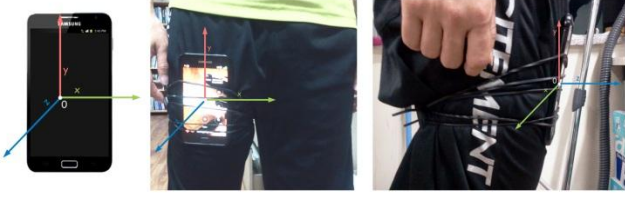downward motion; the Z-axis captures the forward movement.

Fig.1. 3-D coordinate of accelerometer and phone attached to the trouser pocket position

A. Effective Axis Selection

From our observation, the phone does not move horizontally to impact X axis in onward movement. Therefore, we only extract features on Y, and Z axes.

A gait cycle starts with initial contact of the right toe, and will continue until the right toe contacts the ground again. Meanwhile, the association between ground reaction force and forward inertial force together make the Z-axis signal strongly changes and form positive peaks. This associated force also simultaneously forms negative peaks on Y axis when the heel touches the ground. Thus, a gait cycle could be determined by consecutive negative/positive peaks as in figure 2.
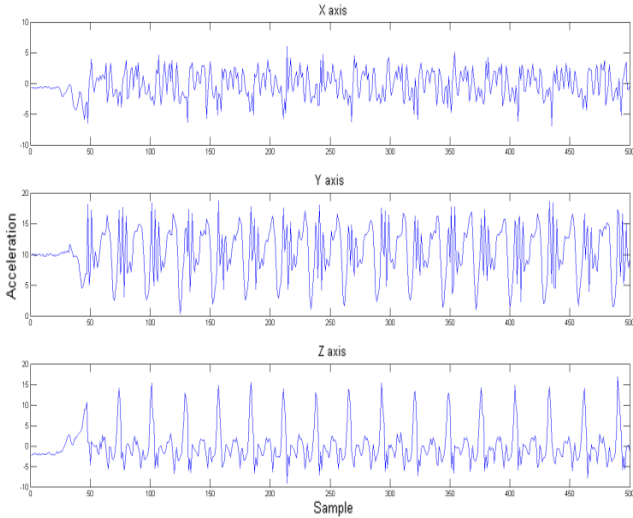

Fig.2. Amplitude in X, Y, and Z axes

From our observation of the activities introduced in this paper scope, we realized that ground reaction force is expressed most clearly especially in jogging and up-stair, down-stair activities by negative peaks. Therefore we consider Y-axis signal as main axis. This work is important contribution because of:

- The first window of a time signal *S(x)* in segmentation is started by first peak on Y-axis instead of choosing first second as first point of the window. This reflects clearly properties of activities since it shows how many peaks are existed in each window. This could enhance accuracy in matching method.

- In overlapping of windows, a next window is still started by a peak which occurred in previous window. Other values on Z-axis are just selected at the same time with Y-axis values. A gait cycle is defined between two consecutive peaks on Y-axis.

B. Linear Interpolation

Due to power saving function and the fact that built-in accelerometer on mobile phone are simpler than standalone sensors, the sampling rate is rather low. Time intervals between two consecutive acceleration values are also not equal. Sensor only outputs value when the forces acting on each axis have a significant change. Therefore, we interpolated the acquired signal to a specific Hz using linear interpolation method to ensure that the time interval between two sample-points will be fixed.

C. Noise Elimination

When accelerometer samples movement data by user walking, some noises will inevitably be collected. These additional noises could have come from various sources (e.g., idle orientation shifts, screen taps, bumps on the road while walking). A digital filter needs to be designed to eliminate noises. A multi-level wavelet decomposition and reconstruction (db6) method was adopted to filter noise.

D. Peak Detection Algorithm

In order to segment window based on Y-axis peaks, we designed a peak detection algorithm as follows:

The original signal is denoted as S(n). First, we extract a set of peaks P from S(n). A data point is called peak if its value is lower than its previous and next one. Let

$$P = \{ d_i \mid d_i < d_{i+1} \wedge d_i < d_{i-1} \} \text{ with i} \in [1 \dots n] \quad (1)$$

where d_i is the ith value in $S(n)$. Threshold T is estimated to determine true peaks using equation (2). The peaks which have magnitudes lower than T are identified as set of true peaks R:

$$T = \mu - k\sigma \quad (2)$$

$$R = \{d_i \in P \mid d_i \leq T \} \quad (3)$$

where μ, σ are mean and standard deviation of all peaks in P respectively and k is the user-defined constant. In our experiment, choosing $k = \frac{1}{3}$ gave the best partition rate.

E. Proposed feature extraction scheme

In this phase, features in our proposed method are separated to different classification features (CF) in time domain and frequency domain. Windows are segmented in time series domain with the fixed 5- second length and 50 percent of overlapping between consecutive windows. All above preprocesses and method for extracting feature are represented in Figure 3.
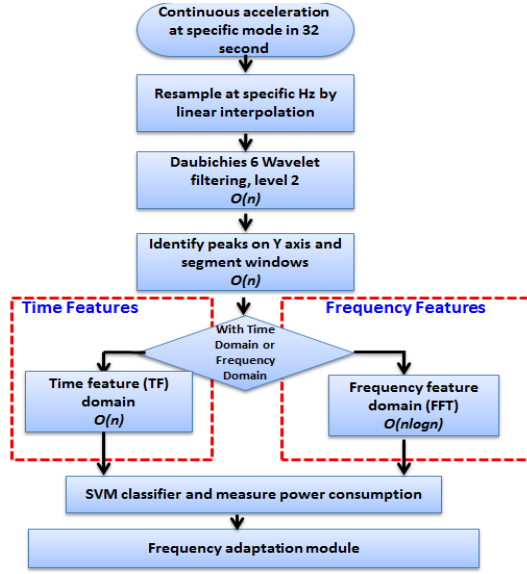
Fig.3. Our proposed method for AR on mobile in pant pocket for benchmarking power consumption at a specific sampling rate

1) SVM using TF method

The following time features (TF) of classifier approach on sliding windows are chosen based on our observation in experiment and recommendations of previous accelerometer based AR research. These features are extracted on effective Y, and Z axes, this is a difference with [3, 4, 6] which used features on all axes:

- Time gap peaks: An average value is computed between two consecutive peaks [7]. This value is only computed on main Y axis.
- Mean and variance acceleration: Mean value is a numerical average of the acceleration values. Variance value shows the mount of variation of the values in the same window [7].
- Accelerometer Energy: This value was also introduced in [2]. Since sampling frequency rate was stable by using linear interpolation, this energy value shows amount of the change on a physical activity. Its value has a significant difference among activities like changes in jogging are occurred in both of Y, and Z axes but the concentration only focuses on Y axis in bicycling. Equation dedicating it in a window size T is presented as:

$$E = \int_{t=t_0}^{t_0 + T} |a_x| dt \qquad (4)$$

where a_x is acceleration at time t on Y, or Z axis.

- Hjorth Mobility and Complexity: In electroencephalography (EEG)[1] signal analysis, Bo Hjorth [8] derived certain features that describe the EEG signal by means of simple time domain since this signal cannot be associated with the since function used in frequency domain. These parameters, namely Activity, Mobility and Complexity, were used to characterize the EEG pattern in terms of amplitude, time scale, and complexity. These values were applied in [9, 10, 2] for emotion assessment, epilepsy diagnosis, and also in accelerometer, respectively. Mobility is a measure of the signal mean

frequency. Complexity measures the deviation of the signal from the sine shape [9]. Both of values are scalar features performed as follows with *var(x)* is variance function of signal *x*, *x'* stands for the derivate of signal *x*:

$$Mobility\ (x) = \sqrt{\frac{var(x')}{var(x)}} \qquad (5)$$

$$Complexity\ (x) = \frac{Mobility(x')}{Mobility(x)} \qquad (6)$$

From analyzed TF features, our classification is acted by using SVM classifier which is used widely in AR [6, 2]. SVM algorithm is set of support vectors which separate training samples to a corresponding class by maximizing margin of hyper-plane among classes. In this work, we use radial basis function (RBF) in order to mapping support vectors to multiple dimension since there have eleven TF attributes.

2) SVM using FFT method

In general frequency domain, amplitudes of fast Fourier transform (FFT) coefficients are considered as effective features like in [7]. For each window, we select the first 40 FFT coefficients on Y, and Z axes respectively. These FFT coefficients on each axis reflect amplitude of basic waves which their combination can reconstruct original signal.

III. EXPERIMENT SETUP AND ANALYSIS ABOUT ACCURACY STANDARD

The methods from above section are firstly applied to SCUTT-NAA data and our data to guarantee the efficiency of our proposed feature extraction.

1) With SCUTT-NAA dataset

SCUTT-NAA dataset [6] contains 1278 samples from 44 subjects. It was collected from ADXL 330 and sampling frequency is 100Hz. Because of limitation of activities in our study, the dataset only provides fully in 31/44 subjects. The Daubechies wavelet decomposition (Db6 at level 3) was used for noise reduction process. Our method uses Y-axis as the main axis for peak detection algorithm. The distance between two true consecutive peaks has to greater than 29 data points. We did not choose overlapping method on sliding windows since there have a large data length in each activity of subjects. In SVM using TF and FFT methods, we separate 512 samples per window. Each window is started from a peak which is the nearest peak to the last peak of previous window. LIBSVM[2] was used to train and predict activities in this dataset and our data.
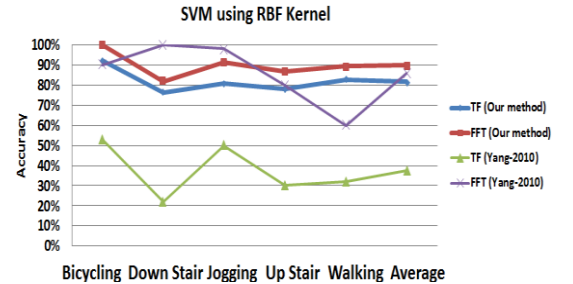

Fig.4. Accuracies for single subject prediction type

---

[1] EEG, http://en.wikipedia.org/wiki/Electroencephalography

[2] LIBSVM, http://www.csie.ntu.edu.tw/~cjlin/libsvm/

We prepared two types from this dataset. In the first type, we separately kept data from every subject for both of training and testing phase. It is called as single prediction type. Data from multiple subjects were mixed in the second type. Figure 4 illustrates the overall accuracy of our methods and Yang [6]. In SVM using RBF graph, our TF method shows a stable accuracy in predicting the activities. Although we and Yang [6] used the same FFT coefficients in SVM classifier, but our method gave more effectiveness because of the significant difference in preprocessing phase. Our contribution in this work is expressed by removing last points in data, selecting effective Y, and Z axes without using X axis, selecting different noise filtering method, and segmenting windows from peaks on main Y axis.

In multiple subject prediction type, we select randomly 60% of data from all subjects to train, and remaining data are used for testing. Figure 5 shows accuracies in different subsets.
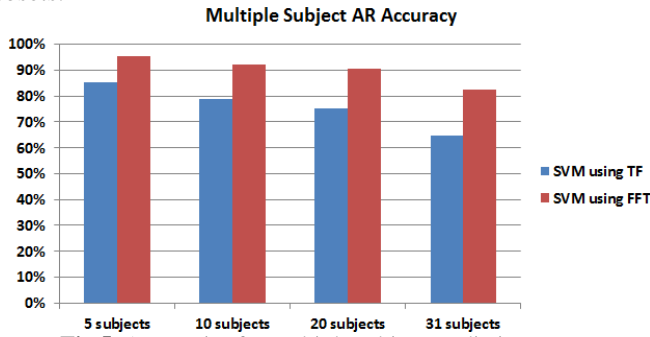

Fig.5. Accuracies for multiple subject prediction type

Our approach produces effectively accuracy in multiple subject prediction type. Accuracies decrease stably through different subsets. Thus, we contributed an improvement in accuracy compared with previous achievement [2] since its authors also used the same dataset.

2) With our self-constructed data

To ensure running successfully in real time environment, we also develop an application for collecting data from tri-axial K3DH accelerometer on Samsung Galaxy Note. It measures acceleration force up to ± 2G. The SFs from the device are in {50 *Hz*, 17*Hz*, 5*Hz*} ranges. We found that classifying and testing at the same frequency have an insignificant effect on the accuracy. Therefore, we performed classifier training at the maximum sampling frequency of 50*Hz*. By this way, the training phase is only performed one time as offline activity for real time classification on user's phone whether another sampling frequency is being activated. The following three plots (Figure 6, 7, 8) describe our observations.
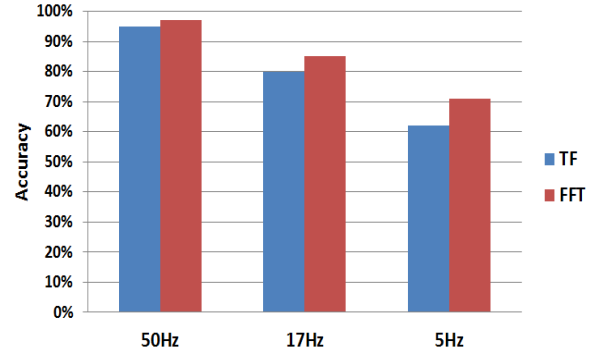

Fig.6. Accuracy at different (SF, CF) combinations

Figure 6 plots the average accuracy of different combinations from five activities performed by four volunteers. This includes two types of classification features and three SFs provided from the accelerometer. As our expectation, (a) the high SFs normally give better predictions, and (b) FFT coefficients perform more effectively classification than TF as in the SCUT-NAA dataset.
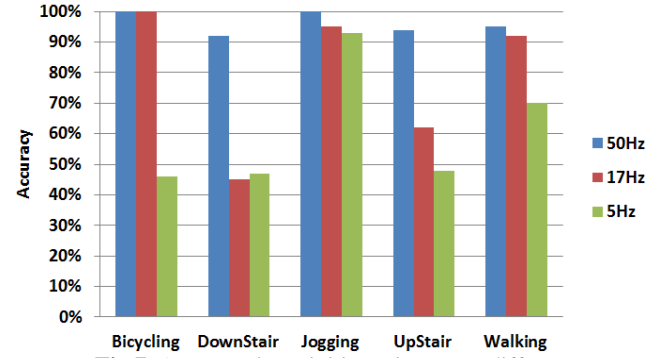

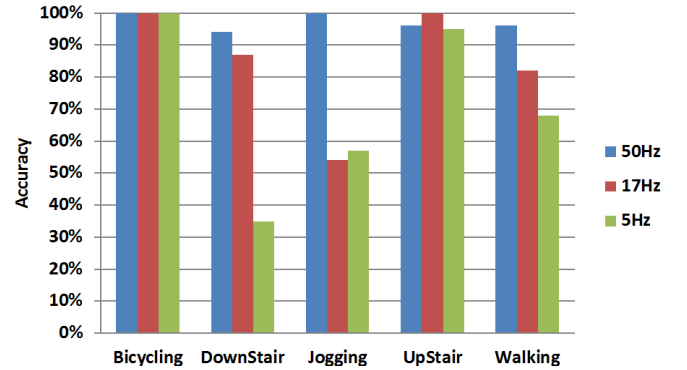**Fig.7.** Accuracy in activities using TF at different Hz

Fig.8. Accuracy in activities using FFT at different Hz

TABLE I.   ACTIVITY RECOGNITION ACCURACY IN DIFFERENT (SF,CF) COMBINATIONS

Activity	Classification Accuracy								
	SF- 50Hz			SF- 17Hz			SF-5Hz		
	TF	*FFT*	*TF+FFT*	*TF*	*FFT*	*TF+FFT*	*TF*	*FFT*	*TF+FFT*
'Bicycling'	1.0	1.0	1.0	1.0	1.0	1.0	0.46	1.0	0.624
'Down Stair'	0.923	0.940	1.0	0.452	0.874	0.53	0.511	0.355	0.584
'Jogging'	1.0	1.0	1.0	0.952	0.542	0.944	0.932	0.57	0.982
'Up Stair'	0.941	0.962	0.942	0.621	1.0	0.74	0.48	0.951	0.524
'Walking'	0.950	0.963	1.0	0.923	0.823	1.0	0.7	0.681	0.812
Power Consumption (%)	**4.8**	**5.6**	**6.0**	**3.8**	**4.1**	**4.5**	**3.2**	**3.4**	**3.8**

From our observation, each individual activity has different effects to (SF, CF) combinations. Accordingly, figure 7 and 8 show the average accuracy in TF, and FFT domains respectively. An effective accuracy of bicycling activity at (17Hz, TF) is retrieved without using the combination of (50Hz, TF). Even with SF = 5Hz, and the use of only FFT features, bicycling activity is also classified correctly. Moreover, 'up-stairs' in (17Hz, FFT) combination gives better accuracy than SF = 50Hz whereas we use the same features. In 'walking' activity, the difference of using combinations at (50Hz, TF) and (17Hz, TF) is not much. It is lower than 3%. Therefore, selecting an appropriate (SF, CF) combination in individual activity should be studied based on energy consumption standard as work in [12].

IV. TRADEOFF BETWEEN ENERGY & ACCURACY

Data analysis in time domain can retrieve useful information which describes characteristics of signal better than amplitude of coefficients in frequency domain [11]. From proposed features in Section II, the computational complexity of feature extraction on time domain is quietly more effective than this complexity on frequency domain:

$$O[TF] = O(n)[Mean] + O(n)[Variance]$$
$$+ O(n)[Energy]$$
$$+ O(n)[Mobility, Complexity]$$
$$+ O(kn)[Peak\ detection]$$
$$with\ k \ll n \quad (7)$$
$$= O(n)$$
$$O[FFT] = O(n log n) \quad (8)$$

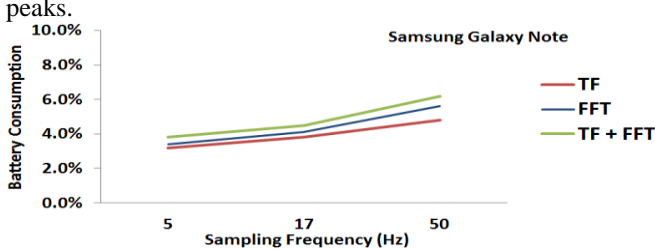where n is total number of data points in signal, k is true peaks.

Fig.9. Average energy consumption for the Samsung Galaxy Note at different (SF, CF) combinations during 0.5 hour

Figure 9 plots the energy consumption (in percentage of full battery charged) over 0.5 hour, with different (SF, CF) combinations. A power measurement module is developed based on battery change event to log these values. Another feature $CF = TF + FFT$ is also supplied, so we can show that it consumes a significant amount of battery over using TF or FFT only. The total energy consumption increases proportionally with SF. The increase in FFT features is not linear since it requires $log n$ function as (8). For all 5 activities, Table 1 shows all possible combinations of (SF, CF) so that we can find out an appropriate selection for individual activity. As 'walking' activity, we can choose (17Hz, TF) which gives an accuracy of 0.923 since it only consumes amount of 3.8%. Even though a (50Hz, TF) combination can give a faintly higher accuracy of 0.95 but its energy consumption reaches significantly amount of 4.8%. Therefore, we defined a strategy in choosing appropriate (SF, CF) combinations as Table 2. It guarantees

effectively both of recognition accuracy and power-saving consumption since we do not need to keep constantly a specific (SF, CF) combination for all activities.

TABLE II.   STRATEGY IN CHOOSING (SF,CF)

Activity	Sampling Frequency	Classification Feature
'Bicycling'	17Hz	TF
'Down Stair'	17Hz	TF
'Jogging'	5Hz	TF
'Up Stair'	5Hz	FFT
'Walking'	17Hz	TF

The *maximum* frequency in these optimal choices is conducted at 17Hz without using the highest SF of the accelerometer at 50Hz. A sequence of different activities is considered as the set of continuous frames which need to be recognized. The important idea is to take a best combination of (SF, CF) for a specific activity that is occurring most commonly in a given frame set. Length of this frame set is denoted by Δ_{Frame}. Based on the continuity of a human activity in fact, the combination is used to recognize next frame set.

Adaptive Strategy in AR on mobile: Since our strategy only contains two types of CF at {TF, FFT} as in Table 2, we generated two classification models including SVM_{TF}, SVM_{FFT} at 50Hz. *Control_Table* contains (SF, CF) combinations of all activity $\{A_1...A_5\}$. *Activity_List* object contains predicted activities by using *classify* function. We used SVM_{TF} as the default classification model in initial phase and *else case* at line 13 since it is more stable than SVM_{FFT} model in Table 1, 2. Δ_{Frame} threshold defines the length of a frame set so that we can select the most popular activity A_i of this period. $\Delta_{Confidence}$ defines a threshold for ensuring that A_i occurs frequently during last frames of *Activity_List* object.

```
1.   load(M ← SVM_TF) ; SF = maximum_SF = 17Hz;
     Control_Table ← [A1:(SF,CF), A2:(SF,CF), …, A5:(SF,CF) ] ;
     t ← 0; Activity_List ← null ;
2.   while Running do
3.       Activity(t) ← classify(Accel_Frame(t), M) ;
4.       add(Activity_List, Activity(t)) ;
5.       if length(Activity_List) ≥ Δ_Frame then
6.           A_i ← max_probability(Activity_List) ;
7.           A_current ← lastActivity(Activity_List, Δ_Confidence) ;
8.           if A_i == A_current then
9.               update(M, Control_Table[A_i]) ;
10.              update(SF, Control_Table[A_i]) ;
11.          else
12.              reset(Activity_List);
13.              load(M ← SVM_TF) ; SF = maximum_SF;
14.          end if
15.      end if
16.      t ← t + 1;
17.  end while
```

V. EXPERIMENT IN ANDROID USERS

Our algorithm is experimented with Android phone users. The study is categorized into three types in which the phone is attached to each user.

- **Non-Adaptive Strategy:** The SF is fixed at 50Hz. We use a classification model with $CF = TF + FFT$
- **Adaptive Strategy:** This type uses our algorithm in order to adjust dynamically (SF,CF)
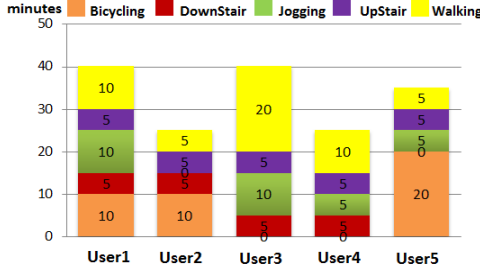- **No AR:** There is no AR application running on phone

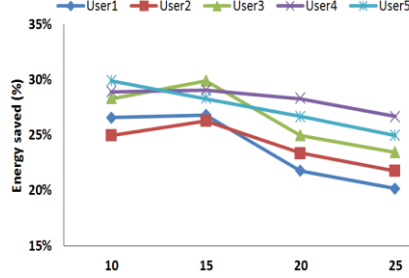**Fig10.** Activities per user during our experiment

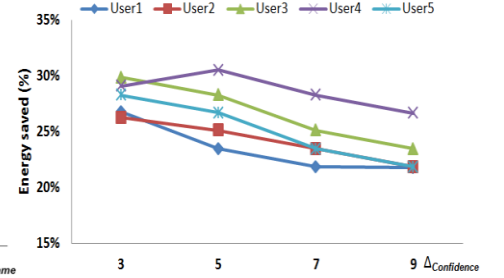**Fig11.** Energy saving on different Δ_{Frame}

Fig12. Energy saving on different $\Delta_{Confidence}$

In order to guarantee accurately our experiment, only data from activities in the scope are collected for recognition. We designed scenarios for five volunteers as shown in Figure 10. Each activity has a specific period per user. We found that the impact of Δ_{Frame} and $\Delta_{Confidence}$ on energy saving changes unstably for all scenarios. Their optimal values are found respectively as in Figure 11, 12. From our observation, $\Delta_{Frame} = 15$ performs the best energy saving since we set $\Delta_{Confidence} = 3$. It can save a value of 28 percentage compared to non-adaptive method ($50Hz$, $TF+FFT$) as (9).

$$Energy\ saved(\%) = 1 - \left[\frac{E(user)}{E(50Hz, TF+FFT)}\right] \qquad (9)$$

where E(user) denotes energy consumption of a (Δ_{Frame}, $\Delta_{Confidence}$) combination in each user's scenario. Figure 12 presents the varying of energy savings of $\Delta_{Confidence}$ when we used $\Delta_{Frame} = 15$. Based on retrieved thresholds, average energy consumptions of three types are presented in Fig 13.
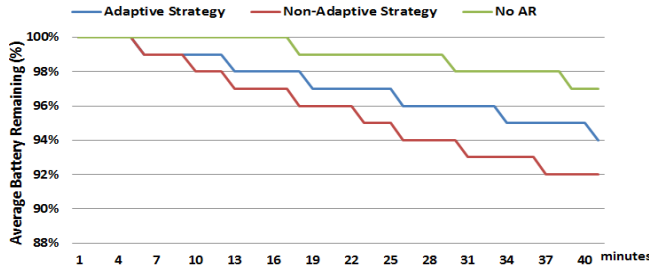

Fig13. Average power consumption in different types

Figure 13 illustrates the average battery drainage time series for all users in our experiment. We can observe easily that adaptive strategy consumes stably battery level through time series. It means that our feature extraction almost predict correct activities in each scenario of all users so that the strategy can transit a known (SF, CF) combination.

VI. CONCLUSION AND FUTURE WORK

Our contribution in AR on mobile accelerometer is expressed through a novel method for feature extraction. It includes peak detection in Y-axis, segmentation method, and power consumption measurement through applying on mobile in real time environment since it guarantees an effective accuracy. A deep analysis in accuracy on a standard dataset and our own data is presented. Finally, we propose an adaptive energy-saving strategy which includes choosing appropriate (SF, CF) for individual activity. Future work might deal with an improvement for our algorithm

based habit of each user in order to optimize Δ_{Frame} and $\Delta_{Confidence}$. Moreover, the optimization in multiple subject prediction type can be studied more based on generating new classification model for a new user without using his/her training data since its training phase could take a lot of time and power battery.

REFERENCES

[1] BBC News-Technology. "Mobile that allows bosses to snoop on staff developed". http://news.bbc.co.uk/2/hi/8559683.stm, 2010-07-27.

[2] M. Khan, S. Ahamed, M. Rahman, R. O. Smith, "A Feature Extraction Method for Realtime Human Activity Recognition on Cell Phones", in Proceedings on Quality of Life Technology, 2011.

[3] S. Das, M. Murphy, "Detecting User Activities using the Accelerometer on Android Smartphones", in TRUST-REU Carnefie Mellon University, 2010.

[4] J. Kwapisz, G. Weiss, S. A. Moore, "Activity Recogntion using Cell Phone Accelerometer", in Proceeding of ACM SIGKDD Conference of Knowledge Discovery and Data Mining, 2010, pp. 74-82.

[5] Frank, B., and Claudia, N. "User Survey on Phone Security and Usage", in Proceedings of the Special Interest Group on Biometrics and Electronic Signatures, 2010, pp. 139-144.

[6] Yang, X., and Lianwen, J. "A Naturalistic 3D Acceleration-based Activity Dataset & Benchmark Evaluations", in Proceeding of IEEE International Conference on Systems, Man, and Cybernetics, 2010.

[7] Bao, L., and Intille, S. "Activity Recognition from User-Annotated Acceleration Data". Lecture Notes Computer Science, 2004, pp. 1-17.

[8] B. Hjorth, "EEG analysis based on time domain properties", in Electroencephalog raphy and Clinical Neurophysiology, 1970, vol.29, no. 3, pp. 306–310.

[9] K. Ansari-Asl, G. Chanel, and T. Pan. "A Channel Selection Method for EEG Classification In Emotion Assessment based on Synchronzation Likelihood", in Proceeding of 15th European Signal Processing Conference (EUSIPCO 2007), 2007, pp. 1241- 1245.

[10] F. S. Bao, and K. J. Oommen. "Automated Epilepsy Diagnosis Using Interictal Scalp EEG", in Proceedings of IEEE Engineering in Medicine and Biology Society, 2009.

[11] Mohammad Omar Derawi, Accelerometer –Based Gait Analysis, A survey, NISK-2010.

[12] Y. Zhixian, and A. Karl. "Energy-Efficient Continous Activity Recognition on Mobile Phones: An Activity-Adaptive Approach", in Proc. of International Symposium on Wearble Computers, 2012.