PREPROCESSING AND CLUSTERING RAW ACCELEROMETER DATA FROM SMARTPHONES FOR HUMAN ACTIVITY RECOGNITION

**Pavel Dinev, Ivo R. Draganov, Ognian L. Boumbarov**

Radiocommunications and Videotechnologies Dept.
Faculty of Telecommunications, Technical University of Sofia
8 Kliment Ohridski Blvd., 1756 Sofia, Bulgaria

e-mail: pr_dinev@abv.bg, idraganov@tu-sofia.bg, olb@tu-sofia.bg

**Darko Brodić**

University of Belgrade, Technical Faculty in Bor
Vojske Jugoslavije 12, 19210 Bor, Serbia

e-mail: dbrodic@tf.bor.ac.rs

***Abstract***

*In this paper is proposed an algorithm for preprocessing and clustering of raw data obtained by accelerometer sensor embedded into a smartphone. It is used by ordinary users while performing sitting, walking and running activities. The goal of the implementation is to enhance the representation of initially generated vectors into compact clusters. As the experimental results reveal it is necessary to introduce an advanced classification approach, such as SVM, in order to recognize the current activity. The method seems promising for application towards users with various medical conditions under remote and prolonged monitoring.*

## 1. INTRODUCTION

Human activity recognition (HAR) becomes increasingly important area of moderm machine intelligence [1, 7, 8]. It includes the ability of automated systems to discriminate various actions performed by humans, such as running, walking, jumping, standing, etc. In the same time the subject is wearing or carrying diferent types of sensors, most often accelerometers. In the recent years buil-in smartphone sensors appeared to be extrmely popular with their low-price and universality wihtin these multipurpose devices [2]. The possible applications include observation of patients with various disabilities in home environment without the presence of medical personnel, monitoring and analysis of actively sporting professionals, investigating the activity performance of military personnel and others [3].

There various methods employed for the recognition of different human activities. In [2-4] Ronao and Cho introduce deep learning neural networks, followed by two-stage continous hidden Markov models, and deep convolutional neural networks. They achieved recognition accuracy up to 95.75%, 96.58% and 99.53% respectively for particular activities.

Committee AdaBoost combining decision trees is proposed by Ugulino et al.[5] where 99.4% is the reported accuracy. In this case wearable sensors are tested during the study including body postures apart from movements.

Siirtola and Röning compared k-nearest neighbour and quadratic discriminant analysis classifiers [6] with a fully embedded implementation on a smartphone adapted to its properties. They announced more than 99% success rate into recognizing the walking activity in real time and as low as close to 70% for other activities from both methods. Offline testing was also done and the rates were considerably lower.

In this paper a basic study is presented over the distribution of the raw accelerometer data – periodic samples along 3 main directions *x*, *y*, and *z* while performing 3 basic activities – running, walking, and sitting. Analyzing the dispersion of the input data allows to find the overlapping of the areas typical for the different activities and to suggest proper classifiaction approaches for further recognition. In Section 2 an algorithm for preprocessing and clustering of the raw data is described, followed by the experimental results obtained with a detailed discussion of what kind of additional processing would be needed and then in Section 4 a conclusion is made.

## 2. ALGORITHM DESCRIPTION

After reading the input raw data a separation is done over the matrix *A* containing the acceleration values over the three dimensions *x*, *y* and *z*. Then a low-pass filtering is implemented over each of the series using a Butterworth filter. It was selected due to its maximally flat frequency response in the passband. The gain *G(ω)* of this filter in the general case of *n*-th order is represented by its transfer function *H(s)*:

$$G^2(\omega)=|H(j\omega)|^2=\frac{G_0^2}{1+\left(\frac{j\omega}{j\omega_c}\right)^{2n}}, \quad (1)$$

where ωc is the cutoff frequency at -3 dB; $G_0$ – the DC gain. The transfer function *H(s)* could be precisely determined by the following product having in mind the properties of the Laplace transform:

$$|H(j\omega)|^2=H(s)H(-s)=\frac{G_0^2}{1+\left(\frac{-s^2}{\omega_c^2}\right)^{n}}. \quad (2)$$

The final form of the system function can also be expressed using the poles:

$$H(s)=\frac{G_0}{\prod_{k=1}^{n}(s-s_k)/\omega_c}. \quad (3)$$

Thus, each vector along the different directions is filtered with the digitized form of this filter into the 1D-signal space. Then, the filtered samples are normalized within the range of [0, 1].

The next stage of the proposed algorithm include clustering of the already preprocessed data using the K-means approach. It is a kind of vector quantization where *n* registered values are split into *k* clusters. Inside each cluster falls a sample which is closest to its mean leading to partitioning of the inputs to Voronoi cells.

Following the steps of the algorithm the aim is to achieve:

$$\arg\min_S \sum_{i=1}^{k}\sum_{x\in S_i}\|x-\mu_i\|^2=\arg\min_S \sum_{i=1}^{k}|S_i|VarS_i \quad (4)$$

where $\mu_i$ is the center from the multitude $S_i$. In other words it is necessary to minimize the squared deviations among adjacent points within one cluster:

$$\arg\min_S \sum_{i=1}^{k}\frac{1}{2|S_i|}\sum_{x,y\in S_i}\|x-y\|^2. \quad (5)$$

Since the cummulative deviation is fixed the inter-cluster squared deviations are going to be maximized as a result.

After the clustering the distribution of the formed groups is analyzed by plotting the scattering of vectors representing the accelerations.

## 3. EXPERIMENTAL RESULTS

Capturing the raw data is accomplished by the use of accelerometer embedded in an Android smartphone. The user can visualize all the parameter changes through application called Accelerometer Monitor. Selecting the option for record the information from the sensor is stored into the phone's memory in the following format:

# Accelerometer Values
# filename: default.txt
# Saving start time: Wed May 13 19:32:17 GMT+01:00 2015
# sensor resolution: 0.038300782m/s^2
#Sensorvendor: Bosch Sensortec, name: BMA250 Acclerometer, type: 1,version : 1, range 39.22
# X value, Y value, Z value, time diff in ms

0.421 3.639 7.776 21
-0.114 3.639 7.967 20
-0.153 3.639 8.236 22
-0.114 3.677 8.427 19
-0.114 3.677 8.427 19
0.306 4.06 8.58 21
0.0 4.367 8.619 21
-0.076 4.405 8.81 19
-1.647 -0.574 9.959 20

# end
#Wed May 13 20:03:39 GMT+01:00 2015

The first three columns are the acceleration values along the *x*, *y* and *z* axes, and the fourth is the time interval between the samples. The graphical user interface of the application is shown in Fig. 1-2.

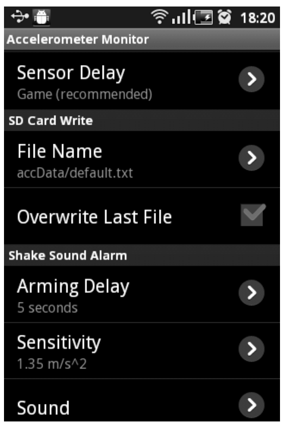
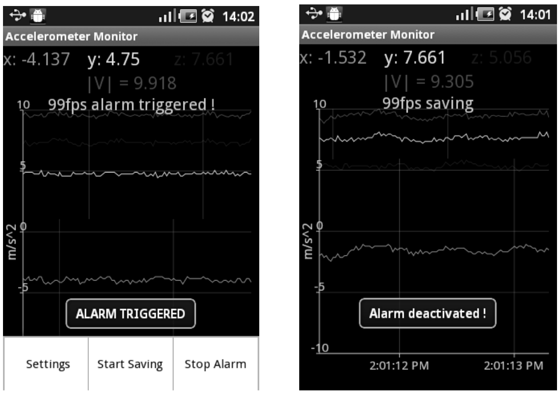



**Figure 1.** Main menu of the Accelerometer Monitor




a) b)

**Figure 2.** Acceleration plots over three dimensions with a) alarm triggered and b) alarm deactivated

After the accumulation of all data a file is preserved for offline processing within the Matlab R2013A environment. It is run over IBM PC compatible computer with Intel CPU 2.8 GHz, 6 GB RAM under the control of MS Windows 7 Ultimate OS.

The source code of the main operations is given below:

```
clc
clear all
close all
fname='C:\Users\Pavel\Desktop\default.txt';% Loading the raw data from file
delim=' ';
A=importdata(fname,delim);

x=A(:,1);%Separating of the raw data matrix into vectors extracting the last fourth column
y=A(:,2);
z=A(:,3);
freq=50;%Sampling rate of the signal coming from the accelerometer
[b a] = butter(50,0.3,'low');% Low-pass filter design
x=filter(b,a,x);%Low-pass filtering for each signal separately
y=filter(b,a,y);
z=filter(b,a,z);
x=(x-min(x))/(max(x)-min(x));%Normalizing the data
y=(y-min(y))/(max(y)-min(y));
z=(z-min(z))/(max(z)-min(z));
A=cat(2,x,y,z);%Concatenating the vectors into a matrix for further clustering
[idx,c] = kmeans(A,3,'distance','cityblock','Replicates',10);% Clustering the data with k-means
figure(1)
hold on
%plot(A(idx==1,1),A(idx==1,2),'r.','MarkerSize',12)
%3-D
%plot(A(idx==2,1),A(idx==2,2),'b.','MarkerSize',12)
% plot
%plot(A(idx==3,1),A(idx==3,2),'g.','MarkerSize',12)
%
plot3(A(idx==1,1),A(idx==1,2),A(idx==1,3),'r.','MarkerSize',12)      %2-D
plot3(A(idx==2,1),A(idx==2,2),A(idx==2,3),'b.','MarkerSize',12)      % plot
plot3(A(idx==3,1),A(idx==3,2),A(idx==3,3),'g.','MarkerSize',12)
plot3(c(:,1),c(:,2),c(:,3),'kx',...!
   'MarkerSize',15,'LineWidth',3)
legend('Running','Walking','Sitting','Centroids',...
     'Location','NW')
title 'Cluster Assignments and Centroids'
grid on;
hold off
```

The experimental results from executing the above code are given in Fig. 3-10. In Fig. 3 the 3-dimensional distribution of the vectors along the main spatial directions is given starting with *x* from the left. It is not symmetrical – the size of the volume occupied by the vectors representing the sitting (in green) is smaller than that of those for running (in red) and inbetween are the vectors for walking (in blue).

As expected, the sitting is characterized with smallest accelerations along all three directions, followed by the walking and highest values are obtained from registering the running. It's also observed that the deviation for the latter is greater varying from very low accelerations for the *y* and *z* directions up to very high ones. Along the *x* axis the effective corss section starts from mid-range magnitudes and goes up to some maximum high above any other activity.

The same behaviour of the vectors' scattering is established from the inverted representation shown in Fig. 4. There it can be seen that the deviation for the walking activity vectors, although a bit lower, is also considerable along the *y* and *z*. The most compact area is occupied by the recorded accelerations for sitting.

When taking into account the possibility of recognizing the different activities it becomes necessary to investigate the separability of the three clusters. In Fig. 5 a projection along *y* is depicted from the 3D distribution. The centers of the clusters are marked with 'x'-s. At this cale it seems that there is good separability between the activities but the closer look suggests the opposite. As Fig. 6 reveals there is overlap with an average value of 0.02 in the normalized range of [0, 1] between sitting and walking. It happens along almost straight line as a boundary with a sharp jump area in the middle. A recommendation of using Support Vector Machine (SVM) into selecting the proper boundaries could be made at this point.

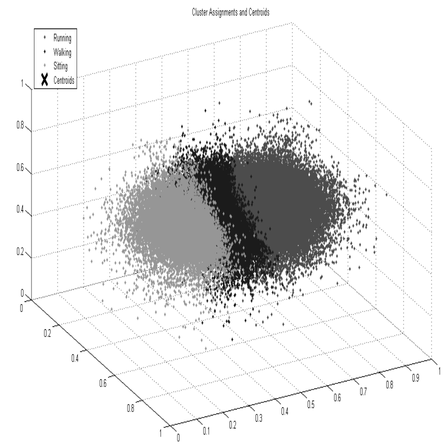



**Figure 3**

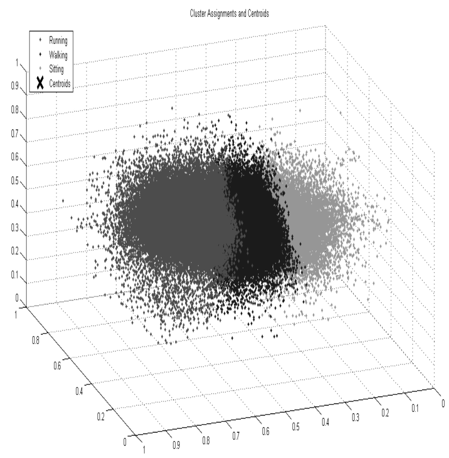



**Figure 4**

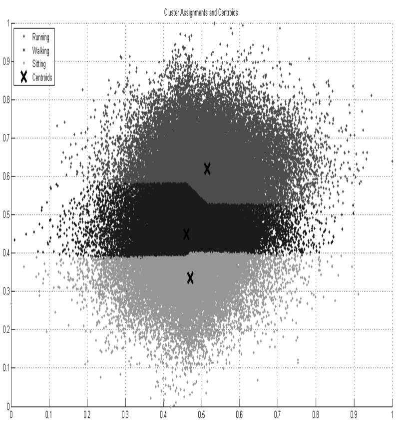



**Figure 5**

Similar is the case along the boundary between the walking and running activities (Fig. 7). Here, the overlapping is 0.04 on average at the same scale with an even higher jump, again around the middle of the border. Once more, SVM could be a proper solution for the case.

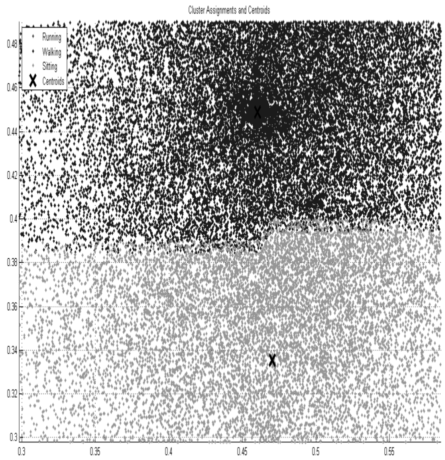



**Figure 6**

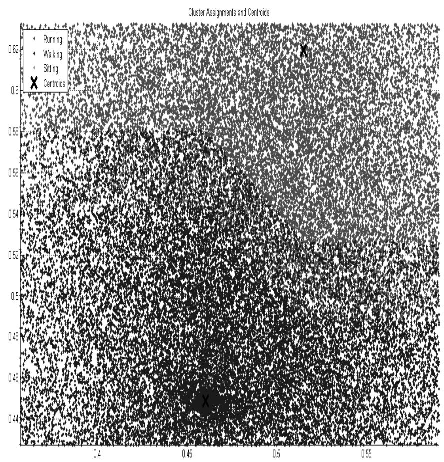



**Figure 7**

In Fig. 8 both the boundaries are visible at greater magnifying level in an area where the jumps are not persisting but still the overlapping is considerable. The diffusion of vectors is less dense between the sitting and walking areas at both ends of the clusters.

Almost the same is the dissipation between running and walking at the lower endbut considerably higher density is observed for these activities at the upper limit, typical for the central parts as well.

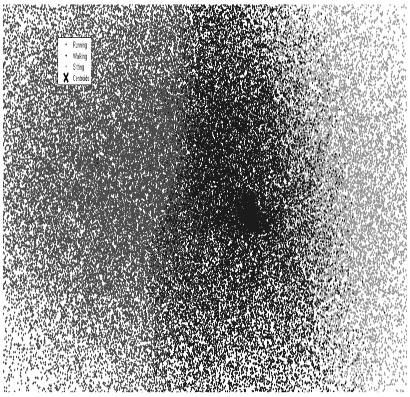



**Figure 8**

## 4. CONCLUSION

The use of accelerometer data for human activity recognition is very convenient because there are quite widespread and cheap sensors. But the processing can be difficult because the signals that are received in the individual activities are highly correlated with each other. There is a need for better classification methods that will inevitably require more time for calculation and more power from the device which can be problematic with mobile ones. All data used in the current study is collected from just one subject. The results are obtained independently.

The application of the proposed approach has its broad potential users, e.g. for elderly or sick people it is possible to calculate the distance travelled or the time spent sitting and being informed that they have not moved enough during the day and have to catch up. It can also be used by active sportspeople who want to keep track of how far they've gone as a distance or how long they have done.

### Acknowledgements

This work was supported by the National Scientific Fund at the Ministry of Education and Science, Republic of Bulgaria, within the project DFNI I02/1 "Intelligent man-machine interface for assistive medical systems in improving the independent living of motor disabled users".

### References

[1] Ghosh, Arindam; Riccardi, Giuseppe. Recognizing human activities from smartphone sensor signals. In: Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014. p. 865-868.

[2] Ronao, Charissa Ann; Cho, Sung-Bae. Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications, 2016, 59: 235-244.

[3] Ronao, Charissa Ann; Cho, Sung-Bae. Human activity recognition using smartphone sensors with two-stage continuous hidden Markov models. In:Natural Computation (ICNC), 2014 10th International Conference on. IEEE, 2014. p. 681-686.

[4] Ronao, Charissa Ann; Cho, Sung-Bae. Deep convolutional neural networks for human activity recognition with smartphone sensors. In: International Conference on Neural Information Processing. Springer, Cham, 2015. p. 46-53.

[5] Ugulino, Wallace, et al. Wearable computing: Accelerometers' data classification of body postures and movements. Advances in Artificial Intelligence-SBIA 2012, 2012, 52-61.

[6] Siirtola, Pekka; Röning, Juha. Recognizing human activities user-independently on smartphones based on accelerometer data. International Journal of Interactive Multimedia and Artificial Intelligence, 2012, 1.5.

[7] Shoaib, Muhammad; Scholten, Hans; Havinga, Paul JM. Towards physical activity recognition using smartphone sensors. In: Ubiquitous Intelligence and Computing, 2013 IEEE 10th International Conference on and 10th International Conference on Autonomic and Trusted Computing (UIC/ATC). IEEE, 2013. p. 80-87.

[8] Su, Xing; Tong, Hanghang; Ji, Ping. Activity recognition with smartphone sensors. Tsinghua Science and Technology, 2014, 19.3: 235-249.