Human Activity Recognition Using Wearable Devices

A thesis submitted in Fulfillment
of the Requirements for the Degree of
Master of Research

by

Jianchao Lu

Student ID: 45041989

Supervisor: Dr. Xi Zheng

MACQUARIE
University

Department of Computing
Faculty of Science
Macquarie University, NSW 2109, Australia

Submitted April 2019

Declaration

I certify that the work in this thesis entitled **"Human Activity Recognition Using Wearable Devices"** has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree to any other university or institution other than Macquarie University. I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signed: .....

Date: .....

Acknowledgements

First of all, I would like to express my sincere appreciation to my supervisor Dr. Xi Zheng for his kindness and patience. He have led me in the correct direction over the last few years not only with his knowledge and experience, but also with thoughtfulness about a young man's personal growth. Literally, without his continuous support and endless guidance, this work would not have been possible. It is my great fortune to have him as my supervisor at Macquarie University. I also would like to thank my colleagues. They have helped me to develop my projects. I wish to express my thanks to Yupeng Jiang and Yao Deng for providing a friendly and enjoyable environment. Many thanks to the staff in the Department of Computing for their administrative help. I would like to thank Sylvian Chow, Melina Chan, Jo Aboud and Karen Leung for their warm support and help. Most important of all, I would like to thank my family. They have always been there for me. Their love, support and encouragement have been the foundation for my life. I wish to thank them for all the opportunities they have made available to me, and for the support they have given me during my life.

Abstract

Human activity recognition (HAR) is a key application on wearable devices in the areas of fitness tracking, healthcare and elder care support. However, inaccurate recognition results may cause an adverse effect on users or even an unpredictable accident. Therefore, it is necessary to improve the accuracy of human activity recognition. This thesis aims to provide effective and efficient HAR methods to address main challenges of HAR, which can be divided into the following three contributions. The first contribution is a novel feature extraction and selection algorithm that addresses the interclass similarity problem in the confounding activity recognition. The second contribution is a novel approach of leveraging local and global features, which addresses both the intraclass variability and interclass similarity problems in HAR. The third contribution is a multiscale feature engineering approach, which leverages local and global features and addresses the negative effect on HAR caused by users' different habits. For the proposed approaches, extensive experiments have been conducted on real datasets or real scenarios. The experiments have demonstrated the proposed methods are superior to the state of the art.

Contents

List of Publications

1. Babin Bhandari, JianChao Lu, Xi Zheng, Sutharshan Rajasegarar, and Chandan Karmakar. Non-invasive sensor based automated smoking activity detection. In *Proc. of EMBC*, pages 845–848. IEEE, 2017

2. Rongjunchen Zhang, Xiao Chen, Jianchao Lu, Sheng Wen, Surya Nepal, and Yang Xiang. Using ai to hack ia: A new stealthy spyware against voice assistance functions in smart phones. *arXiv preprint arXiv:1805.06187*, 2018

3. Jianchao Lu, Jiaxing Wang, Xi Zheng, Chandan Karmakar, and Sutharshan Rajasegarar. Detection of smoking events from confounding activities of daily living. In *Proc. of ACSW*, page 39. ACM, 2019

4. Jianchao Lu, Xi Zheng, Jiaxing Wang, and Jiong Jin. Leveraging local and global features to detect human daily activities. *ACM Transactions on Internet of Things*, 2019, Under Review

5. Jianchao Lu, Xi Zheng, Jiaxing Wang, Wanlei Zhou, and Michael Sheng. MFE-HAR: Multiscale feature engineeringfor human activity recognition using wearable sensors. In *Proc. of MobiQuitous*. ACM, 2019, Under Review

List of Figures

List of Tables

Chapter 1

Introduction

Human activity recognition (HAR) using wearable devices has been widely used in personal use as well as in industries. According to [6], the total market for wearable devices will reach over $170 billion by 2026. With this increasing popularity of wearable devices, it becomes important to assess the feasibility of developing HAR applications for general public based on wearable devices. However, misleading or false reports may cause an adverse effect on users. It could make a bad user experience on wearable fitness tracking system. It may also interfere with clinical decision making if the application poorly discriminates between normal and abnormal users. Worse still, inaccurate recognition results could lead to life-threatening accidents. This happens especially in assisted living for older people if a genuine fall is not detected. Therefore, I need to adopt the technology into any mission critical real-world applications with high accuracy of HAR. However, to achieve satisfactory accuracy in HAR is very challenging as human activities are highly complex and diverse.

1.1 Challenges in Human Activity Recognition

The fundamental challenge in HAR is that human daily life activity is highly complex and diverse which can interfere with the reliability and accuracy of HAR.

1. **Intraclass variability problem in the same activity recognition.**

   In real world scenario, same activity can be performed differently by different

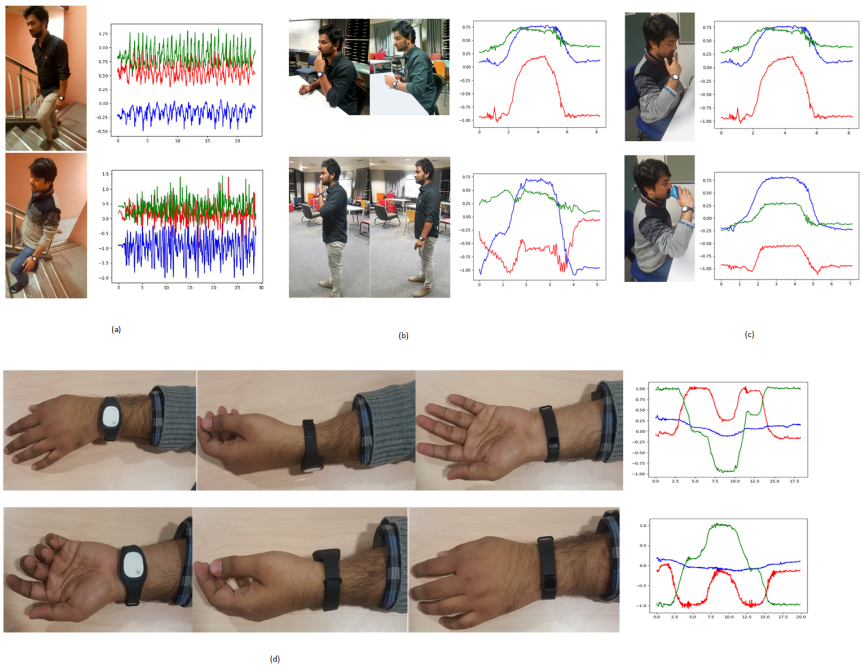


Fig. 1.1. Comparison of the movement patterns

individuals or even by the same person. For example, an activity such as walking upstairs (as shown in Fig.1.1 (a)) is performed by two different individuals. From this figure, it can be seen that the activity remains the same but the performing manner is different. The difference in the change of the signal patterns can be seen when the same activity is performed by two different person. Similarly, smoking style while standing and sitting may be more dynamic (as shown in Fig.1.1 (b)). The change in the signal patterns clearly shows how the same activity varies even if it is performed by the same individual. Thus, in real world, the performing manner of the same set of activity may vary from person to person or even by the same person.

2. **Interclass similarity problem in the confounding activity recognition.**

   Another challenge is to detect those activities that are fundamentally different but show very similar characteristics in the sensor data. For example, activities such as smoking and drinking (as shown in Fig.1.1 (c)) involve similar arm

movement patterns but are completely different set of activities. From the Fig.1.1 (c), it can be seen that the activities are different but the arm movement manners are similar. The signals generated from the arm movement appear similar when both the activities are performed. Such close similarity can often only be resolved by using additional cues captured by different sensors or by analyzing co-occurring activities. Thus, in real world, the performing manner of different set of activities may have similar patterns irrespective of the person.

3. **Different habits cause a negative effect on HAR.**

   Furthermore, different people have their own habits which may cause a significant effect on HAR, contributing to a unstable classification result. The reason is that the sensitivity of sensors' orientation changes. Sensor-based HAR applications are sensitive to some of the sensors' orientation changes. With the lack of orientation independence, users are required to place the sensors, such as the accelerometer and gyroscope, in a specific orientation, which limits their freedom to use these devices. For example, results of the same set of activities vary when the sensors are attached at different body locations. In Fig.1.1 (d) , the same set of activities are performed using a wearable sensor. In the first case, the sensors are attached in front of the wrist whereas in the second case, it is attached to the back of the wrist. The user performs the same activities in both cases but the signal patterns generated are not similar. Thus, the sensors' orientation changes are important factor to consider while developing a real world HAR applications.

1.2 Contributions of the Work

In order to address the above significant and challenging problems in HAR, this thesis makes three major contributions.

1. The first contribution of this thesis is **a novel feature extraction and selection algorithm for confounding activity recognition.**

   (a) To address the interclass similarity problem in the confounding activity recognition, I design orientation features and rotation features, and evaluate the performance of confounding activity recognition using orientation and rotation features, both alone and in combination.

   (b) I propose an ensemble feature selection mechanism for selecting the most relevant features and discarding those irrelevant or redundant, with the main goal of reinforcing the classification accuracy. Four commonly-used feature selection approaches are used to develop the ensemble feature selection machine, which are 'Distance Correlation Coefficient', 'Randomized Lasso', 'Ridge', and 'Recursive Feature Elimination'.

   Experiments conducted on real world datasets illustrate that on average, my methods can recognize the confounding activities with higher accuracy than the existing methods.

2. The second contribution of this thesis is **a novel approach of leveraging local and global features to detect human daily activities.**

   (a) To address both the intraclass variability problem and interclass similarity problem, I propose countable and uncountable activities to better facilitate the understanding of human activities.

   (b) I implement a prototype utilizing a smart phone and a wearable device to collect the data of countable activities.

   (c) I design global and local features and their leveraged feature set for classifying countable and uncountable activities. The key idea is to examine human activities from different perspectives.

   Using one self-collected and another public dataset, my approach is evaluated

to recognize human daily activities with higher accuracy than the state of the art using only one accelerometer.

3. The third contribution of this thesis is **MFE-HAR: a multiscale feature engineering approach for human daily life Activity.**

   (a) I improve my previous work about leveraging local and global features to detect human daily activities. I represent the global features through a comprehensive analysis on acceleration, angular velocity, position, and orientation of body movement; and design the local features based on the characteristics of gesture's orientation.

   (b) I address the problem referring to the different habits cause a negative effect on HAR through extracting features from the Quaternions which are orientation insensitive.

   I conduct an empirical study which shows that MFE-HAR approach is effective for HAR compared with the state of the art.

1.3 Roadmap of the Thesis

This thesis is structured as follows:

- Chapter 2 provides a comprehensive literature review of the state of the art in HAR.
- Chapter 3 proposes a novel feature extraction and selection algorithm for confounding activity recognition. This chapter is based on the published paper [3].
- Chapter 4 proposes a novel approach of leveraging local and global features to detect human daily activities. This chapter is based on the under review paper [4].

- Chapter 5 proposes a multiscale feature engineering approach for human daily life Activity. This chapter is based on the under review paper [5].
- Chapter 6 concludes the work in this thesis and discusses some directions for future research opportunities.

Chapter 2

Literature Review

2.1 HAR Using Multi-device Based Approach

In human activity recognition using multi-Device based Approach, people wear multiple devices simultaneously on different parts of the body. Several works have been done for improving the accuracy of HAR using multi-device based approach. In [7], the authors used five biaxial accelerometer devices worn on different parts of the body for 20 activities recognition. Their experiment result reached an overall mean recognition rate of 80.0%. In their study, they also proposed an additional device on the thigh to increase the mean recognition rate. Similarly in [8], the authors used four wearable devices placed on the wrist, chest, hip and ankle. They recognized activities such as sitting, lying, standing, vacuuming, walking, ascending stairs, and descending stairs. They achieved an overall mean recognition rate of 93.9%. Most of multi-device based approaches achieve impressive results in HAR due to the comprehensive descriptions of an activity from the devices placed on different parts of the body. However, a multi-device based HAR system is not easy to wear on a daily basis. Besides, this approach is usually inefficient in terms of manufacturing, maintenance, and incurring importantly more power usage. Therefore, it is not ideal for real-world applications. In comparison, single-device based approach is more practical to be used in daily life.

2.2 HAR Using Single-device Based Approach

Compared with the multi-device based approach which can obtain comprehensive descriptions of an activity from the wearable devices placed on different parts of the body, single-device based approach has the limitation in terms of the data diversity. Therefore, a few solutions have been developed rely on the single-device based approach to improve HAR accuracy.

2.2.1 Deep Classification Model Approach

Some studies applied deep learning classification models like convolutional neural network (CNN), deep neural network (DNN), and long short-term memory (LSTM) to improve the accuracy of HAR. In study [9], the authors used deep learning approach for human activity recognition using wearable devices. They proposed deep, convolutional and recurrent models, and evaluated their models using three datasets. Their classification results achieved mean f1-score of 90.4% in DNN, 93.7% in CNN, 88.2% in forward LSTM, and 86.8% in bi-directional LSTM. It demonstrated that deep learning approaches outperformed the state of the art classification models. However, these complex models require large dataset for training, and their training costs are often very expensive to incur high computation overhead. Moreover, although the deep classification models give extraordinary predictive abilities, they are more like black box models which make them very difficult to understand and trust. In IoT, especially in the Internet Of Medical Things (IoMT) domain, people need to demystify the complex black-box models and improve transparency and interpretability to make them more trustworthy and reliable so that the medical investigators can understand the main features that affect the outcomes and be able to explain the medical decisions that are made by an decision model. Therefore, deep classification models impose unrealistic constraints on the wearable HAR systems. Since the classification results are determined collectively by the chosen

models, the available dataset and the extracted features [10], I am more leaning towards a feature engineering based approach which does not require complex classification models or large datasets.

2.2.2 Feature Engineering Approach

In the feature engineering approach, people use domain knowledge of human activities to create features that make HAR algorithms work. The features in the sensor data are important to the predictive models people use and will influence the results people are going to achieve. The quality and quantity of the features will have great influence on whether the classification model is good or not. Better features could produce simpler and more flexible models, and they often yield better results. In [11], the authors selected a dataset collected from 10 participants with 13 daily life activities (e.g., walking and jogging) from a smart wristband. They extracted motion features, orientation features and rotation features from accelerometer and gyroscope, respectively, and their feature set achieved highest accuracy of 89% in Random Forest. In another study [12], the authors used accelerometer sensors for various forms of smoking activities detection, such as smoking while sitting and smoking while standing. In their evaluation, 51 candidate features from a wrist device were computed which are largely statistical (e.g., mean, standard deviation, maximum, minimum, median, kurtosis, skew, percentile, SNR and RMS of each window, peak-peak amplitude, peak rate, local peak point, correlation coefficients, crossing rate between axes, slope, MSE and R-squared). Their classification result was impressive with 95% accuracy (F1-score) as they successfully captured the nature of smoking activity. In [13], they used feature selection techniques to improve recognition accuracy. In their study, the authors designed 21 statistical time-domain features and then applied the non-parametric weighted feature extraction (NWFE) combined with the principal component analysis (PCA) to select features to train a LS-SVM classifier. Their experiment showed that the classifier with features selected

by PCA combining NWFE achieved the best performance (99.65% accuracy). A feature-engineering based approach can be an alternative to improve the accuracy of HAR, which does not require a deep classification model, does not require large dataset (which are often not available or inaccessbile), and gives a good understanding about feature importance. As the result, I intend to incorporate the feature engineering with single-device approach into a real-world project to investigate its applicability with more complex scenarios at real-time.

Chapter 3

A Novel Feature Extraction and Selection Algorithm for Confounding Activity Recognition

In my first work [3][1], I propose a novel feature extraction and selection algorithm to address the interclass similarity problem in the confounding activity recognition.

3.1 Confounding Activity Dataset Development

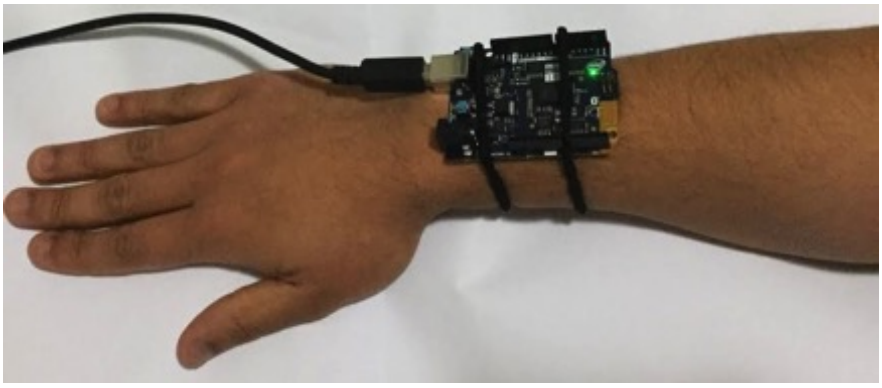


Fig. 3.1. Arduino 101 based sensor device used for capturing data of hand motion related confounding activities.

I design smoking, drinking, eating, scratching head and biting nails as my target confounding activities. During the data collection, five participants (age

[1] This chapter is based on the published paper [3]. The other co-authors are contributing to the writing of the paper while the supervisor is involved with the motivation, design and results discussion of the paper. I am responsible for the solution, implementation and experimentation.

range: 30—35) take part in my data collection experiments and each activity is performed around 5 minutes. All participants perform these activities alone and in a controlled environment. The raw sensor data is collected by using on-board six axis accelerometer and gyroscope from an Arduino 101 development board in the form of a wrist band as shown in Fig.3.1. The accelerometer and gyroscope are ranged at ±2G and ±250 degrees, respectively, with a calibration before data collection. All the raw data are sampled at 20 Hz with minor variations based on on-board clock accuracy.

3.2 Feature Extraction and Selection

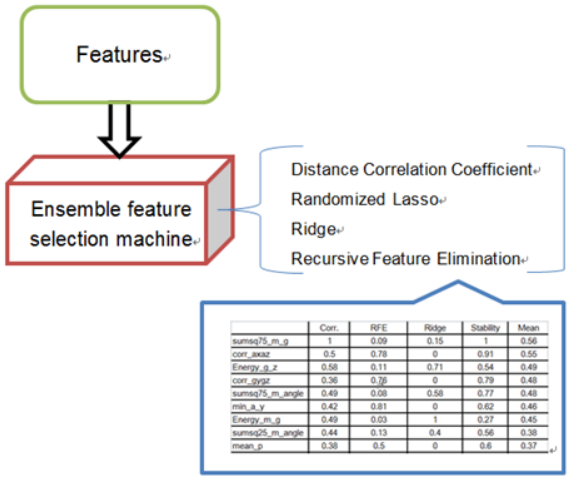
3.2.1 Feature Extraction

In this section, I design three groups of features including **(1) acceleration based orientation features, (2) angular velocity based orientation features and (3) Euler angles based rotation features**. The orientation features are computed from the accelerometer and gyroscope measurements, where the accelerometer is used to measure the orientation with respect to gravity and the gyroscope is used to measure the angular velocity. In addition, I calculate the Euler angles (pitch, yaw, and roll) from the accelerometer and gyroscope values. And then rotation features are calculated from pitch, yaw, and roll, respectively.

The magnitude of acceleration and the magnitude of angular velocity are utilized for the extraction of orientation features, and the magnitude of Euler angles is used for the extraction of rotation features. They are calculated by square-root of the sum of the squares of readings in each accelerometer axis, gyroscope axis, and pitch, yaw, and roll, respectively. Fig.3.2(a) summarizes the features I designed in this work. All these features are extracted under a window size of 30 seconds with 50% window overlapping.

Name	Definition
mean value	$\left(\frac{1}{n}\sum_1^n a_i\right)$
root-mean-square	$\sqrt{\left(\frac{1}{n}\sum_1^n {a_i}^2\right)}$
standard deviation	$\sqrt{\frac{1}{n}\sum_1^n (a_i - mean)^2}$
max	$\max(a_i)$
min	$\min(a_i)$
max-min	$\max(a_i)$-$\min(a_i)$
signal magnitude area	$\frac{1}{t}\left(\int_0^t \lvert a(t)\rvert dt\right)$
correlation	$\frac{cov(x,y)}{\sigma_x \sigma_y}$
energy	$t\sum_{n=0}^{N-1} a^2[n]$
percentiles 25	prctile(a_i, 25)
percentiles 75	prctile(a_i,75)
sum of data below percentile 25	$\sum (a_i)^2 \,;\, a_i < prctile(a_i,25)$
sum of data below percentile 75	$\sum (a_i)^2 \,;\, a_i < prctile(a_i,75)$
interquartile range	prctile(a_i,75) - prctile(a_i, 25)

(a)



(b)

Fig. 3.2. Feature extraction and selection

3.2.2 Feature Selection

A total of 165 features are extracted in this work including 55 orientation accelerometer features, 55 orientation gyroscope features and 55 rotation features. However, not all of them are useful for the confounding activity recognition due to the fact that these features can be relevant (the features that have influence on the evaluation) or irrelevant (the features have no effect on the evaluation). In order to find the relevant feature subset from the entire feature set and increase the efficiency of the machine learning algorithm (prediction), I borrow the idea from ensemble learning algorithm, and then design an ensemble feature selection mechanism for the feature selection as shown in Fig.3.2(b). Four commonly-used feature selection approaches are used in the ensemble feature selection procedure to reinforce the result, which are 'Distance Correlation Coefficient', 'Randomized Lasso', 'Ridge', and 'Recursive Feature Elimination'. After the entire feature set is fed to the ensemble feature selection process, each feature is graded by each of the above four approaches. I scale each feature score to a range from 0 to 1 and then calculate the mean score

for each feature. I sort the mean scores and the unique features corresponding to the highest 15 mean scores are selected as the relevant features.

3.3 Performance Evaluation

For the performance evaluation, a stratified 10 fold cross-validation without shuffling approach is applied. I compute the accuracy, precision, recall and F1-measure for each estimator. In this study, I evaluate the performance obtained from three ways, using only the accelerometer measurements, using only the gyroscope measurements, and using both the accelerometer and gyroscope measurements.

3.3.1 Recognition with Accelerometer

In this section, a total number of 55 orientation features are extracted from the acceleration. Fig.3.3 shows the feature selection result using my feature selection procedure. 18 features are selected from 55 features according to their mean score (unique features of top 15 scores were selected). In Table 3.1, the results with SVM

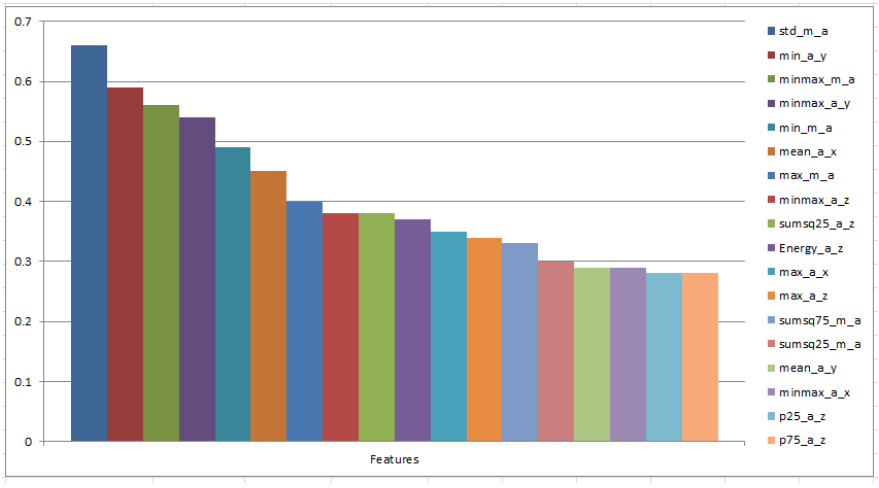


Fig. 3.3. Selected features with mean scores (Accelerometer)

classifier are presented. All activities except eating activity present a low accuracy. The average F1-score for all activities is 63% while F1-scores for each activity are

Table 3.1: Accuracy performance for each activity by SVM (Accelerometer)

Accuracy Performance By SVM			
Activities	Precision	Recall	F1-score
Smoking	0.47	0.5	0.48
Drinking	0.42	0.71	0.53
Eating	0.9	0.82	0.86
Biting Nail	0.62	0.5	0.56
Scratching Head	1.00	0.62	0.77
Average	0.67	0.62	0.63

Table 3.2: Accuracy performance for each activity by RF (Accelerometer)

Accuracy Performance By RF			
Activities	Precision	Recall	F1-score
Smoking	1.00	0.64	0.78
Drinking	0.56	0.71	0.63
Eating	0.92	1.00	0.86
Biting Nail	0.77	1.00	0.87
Scratching Head	1.00	0.88	0.93
Average	0.87	0.84	0.84

48% for smoking, 53% for drinking, 86% for eating, 56% for biting nail and 77% for scratching head. Eating activity achieves the highest score in all of the 5 activities. Table 3.2 presents the results using random forest (RF). Comparing with SVM results given in Table 3.1, the accuracy for all the activities are improved. The average F1-score for all activities using RF classifier is 84% while F1-scores for each activity arse 78% for smoking, 63% for drinking, 96% for eating, 87% for biting nail and 93% for scratching head. In particular, the accuracy for the smoking activity and biting nail activity achieves a significant improvement; up to 30% increase using RF classifier. Again, the performance of eating activity recognition is still the best among these 5 similar activities. According to the results obtained from SVM and RF, although the SVM classifier performs relatively poor, the performance with only accelerometer measurements is still acceptable for these 5 similar activities recognition with an average of 84% accuracy using RF classifier.

3.3.2 Recognition with Gyroscope

In this section, I present the results obtained using only the gyroscope measurements. A total of 55 orientation features are extracted from the gyroscope readings, same as the number of features used with the accelerometer. Fig.3.4 shows the feature selection result using my feature selection procedure. 25 features are selected from 55 features according to their mean score (unique features of top 15 scores were selected). In Table 3.3, the results with SVM classifier are presented. All

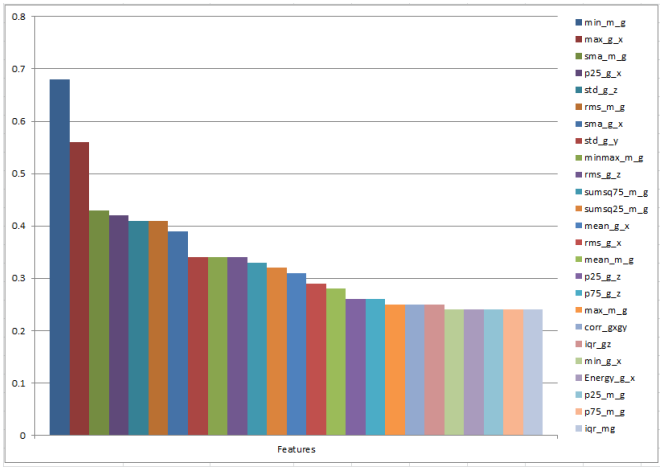


Fig. 3.4. Selected features with mean scores (Gyroscope)

the activities recognized by SVM indicates a less accuracy compared with the SVM results obtained using the accelerometer. The average F1-score for all activities is 33% while F1-scores for each activity are 56% for smoking, 38% for drinking, 58% for eating, 0% for biting nail and 0% for scratching head. Note that none of the biting nail and scratching head activity is recognized by using SVM classifier. Table 3.4 presents the results with RF. Compared with SVM results given in Table 3.3, the accuracy for all the activities are improved. The average F1-score for all activities using RF classifier is 65% while F1-score for each activity is 69% for smoking, 67% for drinking, 70% for eating, 44% for biting nail and 78% for scratching head. Particularly, the accuracy for the biting nails and scratching head activity achieves a significant improvement using RF classifier from 0% to 44% and 78%,

Table 3.3: Accuracy performance for each activity by SVM (Gyroscope)

Accuracy Performance By RF			
Activities	Precision	Recall	F1-score
Smoking	0.43	0.64	0.56
Drinking	0.26	0.57	0.38
Eating	0.83	0.64	0.58
Biting Nail	0.00	0.00	0.00
Scratching Head	0.00	0.00	0.00
Average	0.39	0.40	0.33

Table 3.4: Accuracy performance for each activity by RF (Gyroscope)

Accuracy Performance By RF			
Activities	Precision	Recall	F1-score
Smoking	0.75	0.64	0.69
Drinking	0.55	0.86	0.67
Eating	0.78	0.64	0.70
Biting Nail	0.50	0.40	0.44
Scratching Head	0.70	0.88	0.78
Average	0.67	0.66	0.65

respectively. According to the results represented from SVM and RF, using the gyroscope measurements alone, the SVM classifier still performs poorly for these 5 confounding activities. Although the result of RF classifier is much better than the result of SVM in this section, the performance with only gyroscope solution is still significantly low with an average accuracy of 65%.

3.3.3 Recognition with Both Accelerometer and Gyroscope

Recognition with rotation features only: Firstly, I present the results obtained using only rotation features (features computed from the Euler angles pitch, yaw, and roll). A total of 55 rotation features are extracted from the Euler angles pitch, yaw, and roll. Fig.3.5 shows the feature selection result using my feature selection procedure. 21 features are selected from 55 features according to their mean score (unique features of top 15 scores were selected). In Table 3.5, the

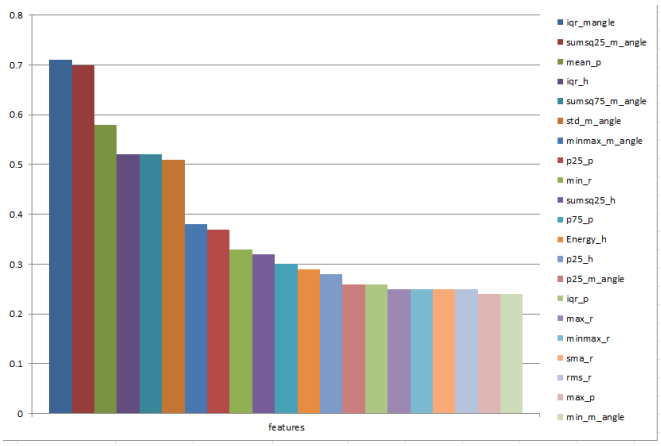


Fig. 3.5. Selected features with mean scores (Rotation features)

Table 3.5: Accuracy performance for each activity by SVM (Rotation features)

Accuracy Performance By SVM			
Activities	Precision	Recall	F1-score
Smoking	0.73	0.57	0.64
Drinking	0.25	0.71	0.37
Eating	0.90	0.82	0.86
Biting Nail	1.00	0.50	0.67
Scratching Head	1.00	0.00	0.00
Average	0.64	0.90	0.55

results of SVM classifier are presented. All the activities recognized by SVM show worse accuracy compared with the SVM results obtained using orientation features from accelerometer, but better than the SVM results obtained using the orientation features from gyroscope. The average F1-score for all activities is 55% while F1-scores for each activity are 64% for smoking, 37% for drinking, 86% for eating, 67% for biting nail and 0% for scratching head. Note that the scratching head activity is none recognized by the SVM classifier. Table 3.6 presents the results with RF. Compared with SVM results given in Table 3.5, the accuracy for all the activities are improved. The average F1-score for all activities using RF classifier is 80% while F1-scores for each activity are 71% for smoking, 93% for drinking, 100% for eating, 95% for biting nail and 38% for scratching head. Particularly, the recognition

Table 3.6: Accuracy performance for each activity by RF (Rotation features)

Accuracy Performance By RF			
Activities	Precision	Recall	F1-score
Smoking	0.71	0.71	0.71
Drinking	0.88	1.00	0.93
Eating	1.00	1.00	1.00
Biting Nail	1.00	0.90	0.95
Scratching Head	0.38	0.38	0.38
Average	0.80	0.80	0.80

of drinking, eating and biting nails activities achieves a high accuracy using RF classifier. According to the results represented from SVM and RF using rotation features alone, the SVM classifier still shows a poor performance in recognizing these 5 confounding activities. However, the performance with using rotation features only can still be acceptable for these 5 confounding activities with average accuracy of 80% using RF classifier.

Recognition with orientation and rotation features: Then,I present the results obtained by using both orientation and rotation features. A total of 165 features are extracted from the accelerometer and gyroscope readings (55 orientation accelerometer features, 55 orientation gyroscope features and 55 rotation features). Fig.3.6 illustrates the feature selection result using my feature selection procedure. 19 features were selected from 165 features according to their mean score (unique features of top 15 scores were selected). In Table 3.7, the results with SVM classifier are presented. In this section, the SVM classifier makes a superior performance compared with the performance in Section 3.3.1, Section 3.3.2 and 'Recognition with rotation features only' part in Section 3.3.3. The average F1-score for all activities achieves 78% while F1-scores for each activity are 90% for smoking, 56% for drinking, 80% for eating, 75% for biting nail and 80% for scratching head. Smoking activity achieves the highest score in all of the 5 activities. Table 3.8 presents the results with RF (RF). Compared with SVM results given in Table 3.7, the

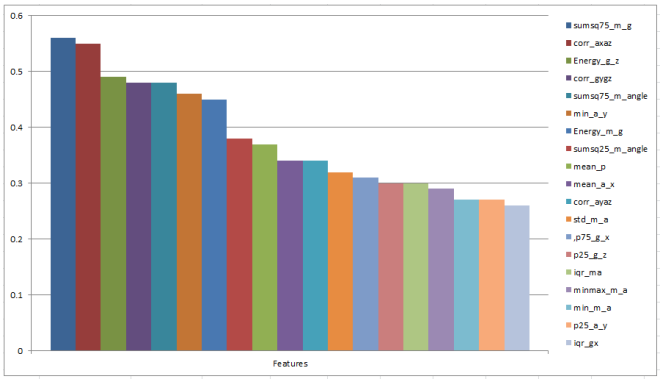


Fig. 3.6. Selected features with mean scores (Orientation and Rotation features)

Table 3.7: Accuracy performance for each activity by SVM (Orientation and Rotation features)

Accuracy Performance By SVM			
Activities	Precision	Recall	F1-score
Smoking	0.82	1.00	0.90
Drinking	0.45	0.71	0.56
Eating	0.89	0.73	0.80
Biting Nail	1.00	0.60	0.75
Scratching Head	0.86	0.75	0.80
Average	0.83	0.78	0.78

Table 3.8: Accuracy performance for each activity by RF (Orientation and Rotation features)

Accuracy Performance By RF			
Activities	Precision	Recall	F1-score
Smoking	0.93	1.00	0.97
Drinking	0.67	0.86	0.75
Eating	1.00	0.64	0.78
Biting Nail	0.91	1.00	0.95
Scratching Head	0.88	0.88	0.88
Average	0.90	0.88	0.88

accuracy for all the activities are improved. The average F1-score for all activities using RF classifier is 88% while F1-scores for each activity are 97% for smoking, 75% for drinking, 78% for eating, 95% for biting nail and 88% for scratching head. Again, the performance of smoking activity recognition is the best among these 5 similar activities. Furthermore, for the RF classifier, the average F1-score of 88% is the highest score compared with the 84% in Section 3.3.1, 65% in Section 3.3.2 and 80% in 'Recognition with rotation features only' part of Section 3.3.3. According to the results represented from SVM and RF, although the average accuracy of 78% using SVM classifier is not as remarkable as the average accuracy of 88% achieved using RF, the performance with orientation and rotation features are reasonably high for these 5 similar activities using both the classifiers.

3.4 Summary

The first thing that can be noted from this work is, the solution of using only one accelerometer sensor to recognize the 5 highly related activities, namely, smoking, drinking, eating, biting nail, and scratching head, can perform in a similar way using both an accelerometer and gyroscope with RF classifier. However, adopting the solution of using both the accelerometer and gyroscope features can improve the prediction result slightly using RF classifier, or strongly using SVM (see the F1-score in Table 3.7). In contrast, the solution of using only gyroscope features without any other sensor measurements cannot provide good performance in detecting these activities. The second thing that I observed from my research is, although the performance of using the gyroscope features alone to predict these 5 highly related activities is not satisfactory, however, as shown in Fig.3.8, these features extracted from the gyroscope could be highly relevant to the prediction. Fig.3.8 presents 19 features selected in Fig.3.6, in which 7 out of 19 features are contributed by the gyroscope, accounts for 36.80% of the whole feature set, and 3 features

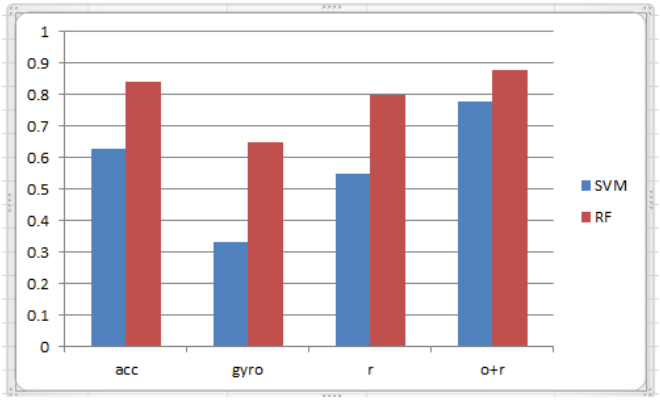


Fig. 3.7. F1-score summary under different solutions.
Legend acc: Accelerometer only features, gyro: Gyroscope only features, r: Rotation feature, o+r: Orientation and Rotation features

are contributed by the accelerometer and gyroscope combination, accounts for 15.8%. More than 50% of the features are directly or indirectly related to gyroscope features.

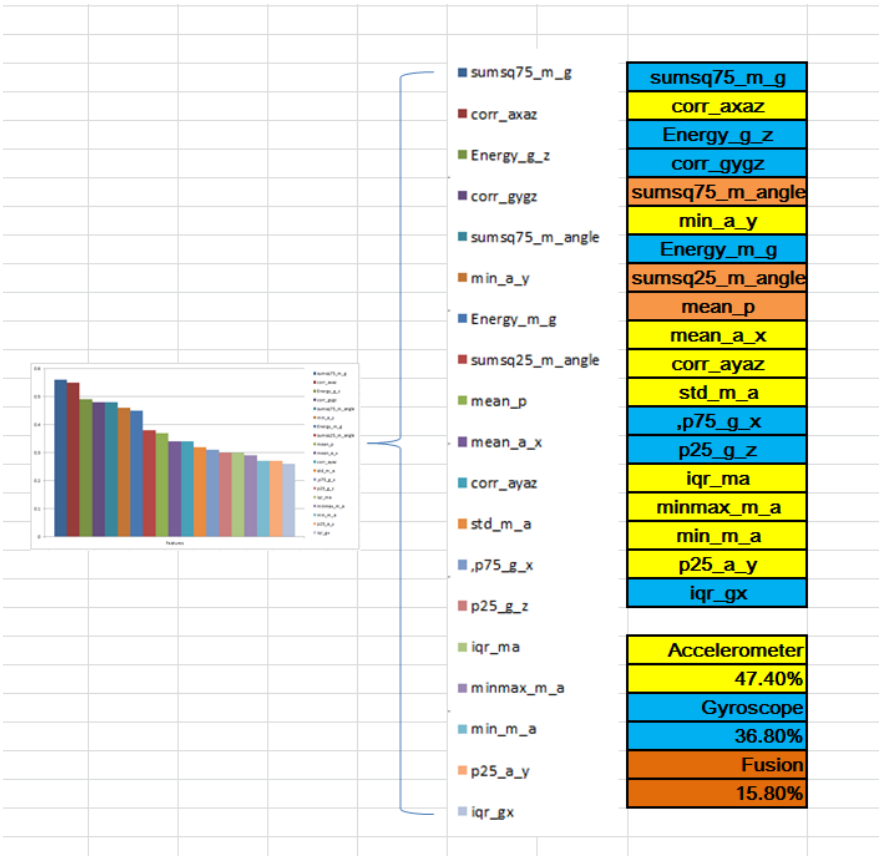


Fig. 3.8. Percentage of features occupied from an Accelerometer and Gyroscope

Chapter 4

Leveraging Local and Global Features to Detect Human Daily Activities

This study[4] [1] continues and expands my previous work [3] from confounding activities recognition to human daily activity recognition. It addresses both the intraclass variability problem and interclass similarity problem by leveraging local and global features to detect human daily life Activity.

4.1 Important Definition of Countable and Uncountable Daily Life Activities

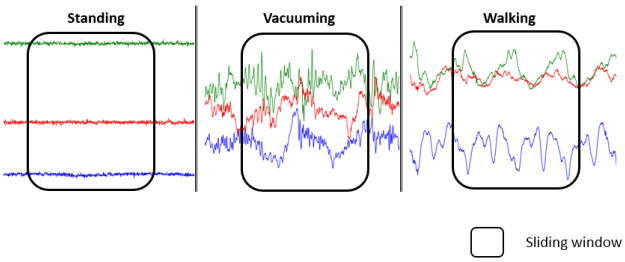
Depending on human daily life activities complexity, I categorize them into two different levels: **'gestures'** and **'activities'**. **'Gestures'** are elementary movements of a human's body part, and are the atomic components describing the fundamental human behaviors, such as 'lifting an arm' and 'raising a leg'. **'Activities'** are composed of multiple gestures, which can describe many complex human behaviors, for instance, 'walking' and 'smoking'. Based on these two categories, I then define the **'uncountable activity'**, which usually is a long-duration activity and composed by countless gestures or a single long-duration gesture. The reason I call this kind

[1] This chapter is based on the under review paper [4]. The other co-authors are contributing to the writing of the paper while the supervisor is involved with the motivation, design and results discussion of the paper. I am responsible for the solution, implementation and experimentation.

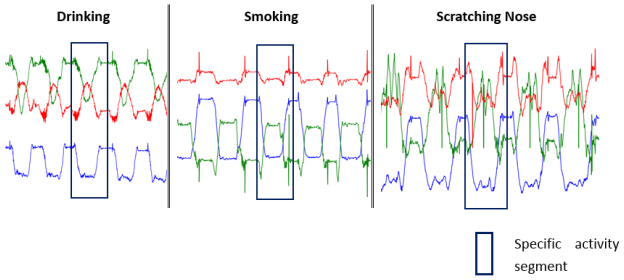
of activities as uncountable activities is that they are difficult to be divided into a certain number of gestures. For example, a 'walking' activity includes innumerable steps (gestures) and it is hard to define a specific number of steps to represent a 'walking'. On the other hand, I define the **'countable activity'** as a short-duration activity and composed by countable gestures, since a countable activity can easily be divided into a certain number of gestures. For instance, smoking activity comprises a set of short arm movements (gestures), which is lift arm, hold-still, puff, and lift down. Compared with the classification of countable activities, the classification of uncountable activities is relatively easy. The reason is that uncountable activities are often processed using a common sliding window method [7, 8] with a given window size and the signals of uncountable activities usually have identifiable patterns as shown in Fig.4.1(a). However, processing the signals of countable activities are often required to design a specific activity segmentation algorithm[11] and usually these signals have fairly similar patterns which increase the difficulty in the classification as shown in Fig.4.1(b).

4.2 Leveraging Local and Global Features to Detect Human Daily Activities

In this work, I extract my features from a global perspective (global feature) and a local perspective (local feature), respectively. Global features describe an overall perspective of the activities as they illustrate the general trend of movement in an activity, while local features are supplementary to global features which provide more specific details of an activity in a subtle time period of movement. The key idea is to examine the daily life activity from different perspectives, and attempt to give a comprehensive description about characteristics of each activity through the combination of the global and local features. An overview of my approach for activity detection is shown in Fig.4.2. Firstly, the raw data are collected from



(a) Uncountable activities of standing, vacuuming and walking



(b) Countable activities of drinking, smoking, and scratching nose

Fig. 4.1. Example of uncountable activities and countable activities signal patterns

the wearable sensors, by performing various activities of real life. These signals are then preprocessed, which include noise removal, signal segmentation and resampling procedures. These operations are important in order for the subsequent operations, such as the feature extraction, to perform efficiently on these real data. I then perform the feature extraction from these signals in order to better facilitate the extraction of activities accurately. In this process, I propose a novel feature extraction procedure, consisting of two sets of features, namely, global and local features. These local and global features aid the detection of uncountable and countable activities accurately. Subsequent to feature extraction, feature selection is performed to identify the most useful and relevant features for HAR.

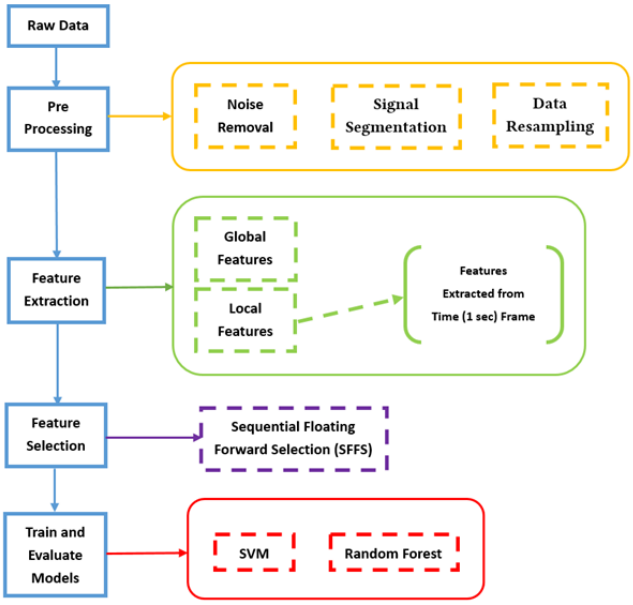


Fig. 4.2. Data processing pipeline for activity detection

4.2.1 Data Acquisition

In this work, one public dataset and one self-collected dataset are used for the evaluation of my feature extraction approach.

Daily Life Activities (DaLiAc) Dataset [8]. I select DaLiAc as the uncountable activity dataset. The authors of DaLiAc designed 13 daily life activities including sitting, lying, standing, washing dishes, vacuuming, sweeping, walking, ascending stairs, descending stairs, treadmill running, bicycling on ergometer, and rope jumping. In this work, for the purpose of making a benchmark study, I select sitting, lying, standing, vacuuming, walking, ascending stairs, and descending stairs as my target uncountable activities.

Self-Collected Arm-Movement activities (AmA) Dataset. In this study, I redesign 6 confounding activities consisting of arm movements, namely, smoking, drinking, eating, scratching head, scratching nose, and using phone, which form my countable activities dataset. I remove 'biting nail' from my dataset due to the difficulty of its collection in the real world. These activities mimic the hand motion of

a person in real life setting with an uncontrolled environment, which is a improved method of data collection compared with the one I used in the previous study[3] under a controlled environment. AmA dataset is collected from 21 participants (each participant performed 6 activities) using on-board three axis accelerometer built on an Arduino 101 development board, in the form of a wrist band prototype as shown in Fig.4.3. These three-axis accelerometers have a range of ±2G, and the raw data are sampled at 50 Hz.

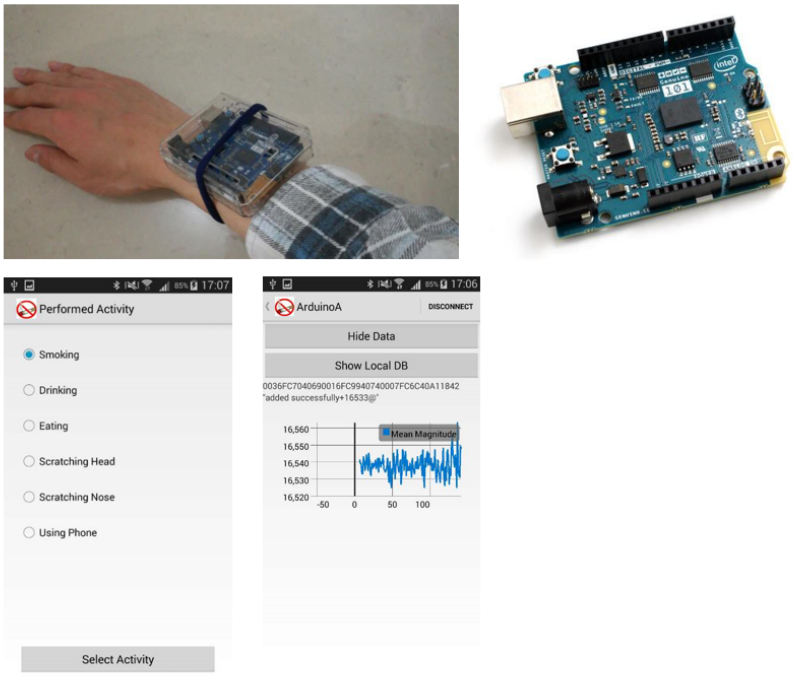


Fig. 4.3. Arduino 101 based wrist band prototype used for capturing hand motion related activities

4.2.2 Noise Removal

Raw signals usually contain noise that arises from different sources, such as sensor miscalibration, sensor errors, errors in sensor placement, or noisy environments[14]. These noisy signals adversely affect the signal segmentation, feature extraction and then significantly hamper activity prediction. Since the gravitational force is

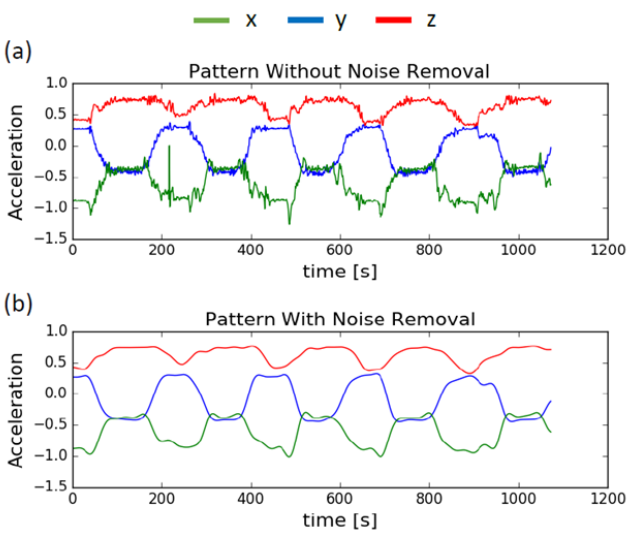


Fig. 4.4. Accelerometer signals of example smoking activity with (a) and without (b) noise removal

assumed to have only low frequency components, I use fourth order Butterworth low-pass filter for the noise removal. Fig.4.4 compares the sensor signal pattern with (Fig.4.4(a)) and without (Fig.4.4(b)) noise removal for the same activity.

4.2.3 Data Segmentation

Sliding window for DaLiAc. For the DaLiAc data, I use a sliding window approach for further processing. I select the window size with five seconds, as used in [8].

Specific activity segmentation for AmA. The activity data collected in the AmA dataset, such as for smoking and drinking, comprise raw accelerometer sequences, in addition to non-activity periods (stationary signals) in between the activities. In order to retrieve the signals corresponding to a particular activity (excluding non-stationary signals) from the sensor streams, I divide the raw accelerometer signals into a set of individual segments, where each segment corresponds to a 'specific' activity. Previously, many segmentation methods have been used in various activity detection studies, like top-down [15], bottom-up [16], sliding window [17], and Sliding Window And Bottom-up (SWAB) [16] methods. In this work, I use the

bottom-up based segmentation method for processing the AmA data.

Further segmentation for DaLiAc and AmA. Moreover, I develop another kind of the frame, which is based on each sliding window or individual 'specific' activity segments in my experiment. I further divide each window, and the 'specific' activity segment, into 4 frames, where each frame includes one second data of an activity as shown in Fig.4.5.

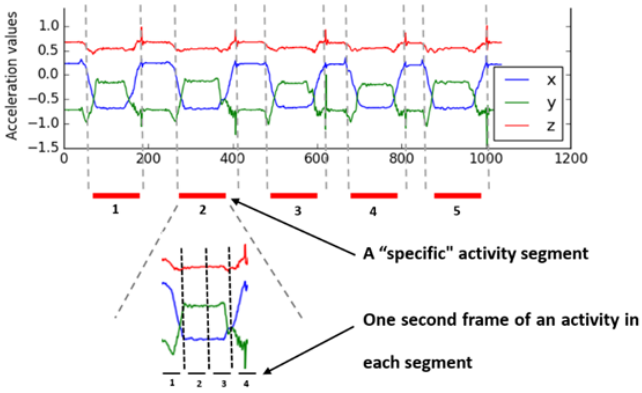


Fig. 4.5. Example of a further data segmentation from an activity segment

4.2.4 Data Resampling

Data Resampling for DaLiAc. In this section, I use the data from DaLiAc dataset without resampling, as used in the study [8].

Data Resampling for AmA. From the collected data in AmA, I observe that all the extracted segments are not of the same length, since all the activities are performed by different subjects mimicking the real life activity. In this study, I aim to extract local features from activity segment besides global features. In order to make the local features dimension consistent across different activities and subjects, all the segments need to be adjusted to have the same length. Therefore, I re-sample these segments to form a uniform length of samples across all the segments. In my experiment, as the minimum number of samples found in segments is 115 (2.3 seconds), whilst the maximum number of samples found in segments is 235 (4.7

seconds), I choose a uniform length of 200 samples corresponding to 4 seconds for each 'specific' activity segment.

4.2.5 Feature Extraction

Fig.4.6 illustrates my proposed framework of feature extraction approach , where the features that are extracted from the sliding window or complete activity segment forms the 'global features', and these extracted from a sub time frame of the sliding window or complete activity segment forms the 'local features'. I calculate time domain features and (or) frequency domain features to describe an overall perspective of the activities in global features. In addition, I calculate time domain features and (or) frequency domain features for each sub time frame for the purpose of providing more specific details of an activity in a subtle time period of movement. Furthermore, I calculate the Delta (△) features between two consecutive sub time frames for indicating local level variations within an activity.

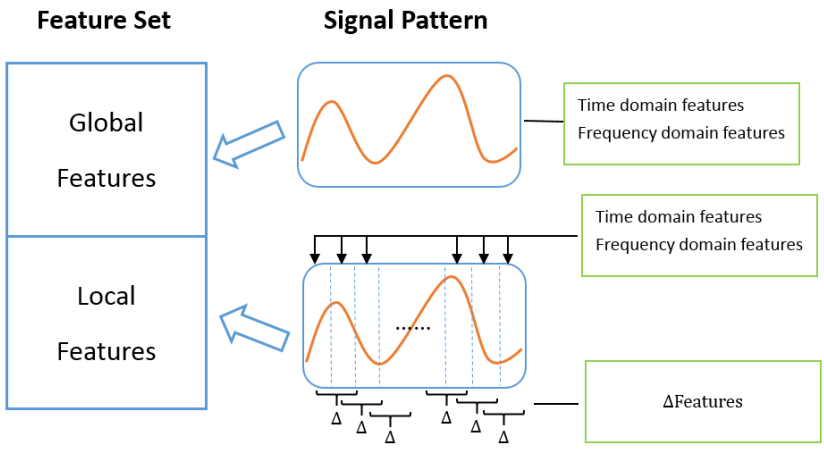


Fig. 4.6. Proposed feature extraction approach framework

In [8], six generic features are computed for each sliding window in each of the three accelerometers and gyroscope axes. In my work, I extract six features as global features for each sliding window and 'specific' activity segment (which excludes

Table 4.1: List of global features computed for each axis

1	minimum	4	variance
2	maximum	5	skewness
3	mean	6	kurtosis

non-stationary signals) from every accelerometer axis. Besides the three-axis (x, y and z) accelerometer signals, the resultant Magnitude of x, y and z axis are also used for feature extraction.

The six global features are listed in Table 4.1. In total, this results in 24 global features per sensor node (by timing x, y, z, and Magnitude).

Similar to global features, mean, variance, maximum, minimum, skewness and kurtosis are extracted for each frame to consider them as local features. Thus, they comprise 24×5 (frames) = 120 local features for DaLiAc dataset, and 24×4 (frames) = 96 local features for AmA dataset. Moreover, 3 Delta local features, corresponding to the changes of signal pattern distribution are also calculated for this study.

Before explaining these 3 Delta local features, I define the parameters that are used to represent a specific activity segment, and the framing approach to extract these new local features.

Definition 1 *Raw Accelerometer Stream A records acceleration in each axis, given as,* $A = (a_x, a_y, a_z, t)$ *is a tuple with accelerations (x; y; z) and timestamp (t).*

Definition 2 *The axis' value of A for 'specific' activity segments is given as* $\boldsymbol{A_{axis,i}} = \{a_{axis,i}\}_{i=1}^{m}$*, where* $axis = \{x, y, z, mag\}$ *and* $i = \{1, 2, 3, \ldots, m\}$*, which represents the* i_{th} *"specific" activity segment in A.*

Definition 3 *Accelerometer frame* $\boldsymbol{A^{(f)}_{axis,i}}$ *is a set of 'single frame length' (in this case* 1 *second length) accelerometer records in each axis for 'specific' activity segments, given as,* $\boldsymbol{A^{(f)}_{axis,i}} = \{a_{axis,i}((f-1)*n+1 : f*n)\}_{i=1}^{m}$*, where* $n = 50$ *is the number of samples in a 'single frame length' accelerometer record,* $f = \{1, 2, 3, 4\}$ *represents the* f_{th} *'single frame length' accelerometer record.*

My proposed 3 Delta local features are listed below:

1. **Change of Correlation of Acceleration Values (CCA).** I calculate the Pearson correlation coefficient between a pair of accelerometer axes for each window frame of activity using:

$$CA^{(f)}_{\{a,b\}} = \frac{\sum(A^{(f)}_a - A^{-(f)}_a)(A^{(f)}_b - A^{-(f)}_b)}{\sqrt{\sum(A^{(f)}_a - A^{-(f)}_a)^2 (A^{(f)}_b - A^{-(f)}_b)^2}} \tag{4.1}$$

   where $\{a,b\} \in \{\{x,y\},\{x,z\},\{y,z\}\}$. Then I compute the change of Pearson correlation values between two consecutive frame lengths (seconds) in each segment using:

$$CCA^{(f)}_{\{a,b\}} = CA^{(f+1)}_{\{a,b\}} - CA^{(f)}_{\{a,b\}} \tag{4.2}$$

   where $\{a,b\} \in \{\{x,y\},\{x,z\},\{y,z\}\}$. A value of CCA closer to zero indicates a small change in the correlation values of acceleration in two consecutive frames.

2. **Change of Similarity of Acceleration Values (CSA).** I use fast dynamic time warping (FastDTW) algorithm to determine the similarity between a pair of accelerometer axes for each frame. I use FastDTW over the traditional dynamic time warping (DTW) since FastDTW is an accurate approximation of DTW which incurs lower time and space complexity [18]. First, I calculate the similarity of acceleration (SA) by:

$$SA^{(f)}_{\{a,b\}} = fastDTW(A^{(f)}_a, A^{(f)}_b) \tag{4.3}$$

   where $\{a,b\} \in \{\{x,y\},\{x,z\},\{y,z\}\}$. Then I compute CSA between two consecutive frames in each segment using:

$$CSA^{(f)}_{\{a,b\}} = SA^{(f+1)}_{\{a,b\}} - SA^{(f)}_{\{a,b\}} \tag{4.4}$$

   where $\{a,b\} \in \{\{x,y\},\{x,z\},\{y,z\}\}$. A value of CSA closer to zero indicates a small change in SA between two consecutive frames.

3. **Difference of Linear Approximation Error of Acceleration Values (△LAEA)** I calculate the best fitting line for each axis and magnitude signal of each frame using least-squares approximation, and compute their difference of linear approximation error between two consecutive frames using:

$$\Delta LAEA_i^{(f)} = lstsqE(A_i^{(f+1)}, A_i^{(f)}) \tag{4.5}$$

where $i \in \{x, y, z, mag\}$ and $lstsqE(\cdot)$ is the function for calculating the least-squares error, $\Delta LAEA^{(f)}$ represents the difference of linear approximation error in each axis or magnitude signal between two consecutive frames $f+1$ and f. A value of $\Delta LAEA$ closer to zero indicates similar goodness of fit for consecutive frames.

These new features comprise $(3+3+4) \times 4(\Delta frame) = 40$ local features. Hence, a total of $96 + 40 = 136$ local features are extracted for this work.

Finally, I select an optimal feature subset from the whole feature set using sequential floating forward selection (SFFS) algorithm [19]. As SFFS is a wrapper method, I wrap K-Nearest-Neighbor (KNN) classifier with K = 5 and 20-fold cross-validation for the feature selection.

4.3 Evaluation and Discuss

4.3.1 Evaluation Methods

Using the two datasets explained previously (i.e., DaLiAc and AmA datasets), I compare the state of the art solutions in detecting HAR (i.e., detecting only uncountable activities using global features) with my proposed approach to detect both uncountable and countable activities leveraging global and newly introduced local features. I implement my algorithm in python 2.71, using Scikit-learn 0.17.1 package which is an open source Python library to implement machine learning algorithms [20], and Mlxtend (machine learning extensions) which is a Python library of useful tools

for the day-to-day data science tasks [19]. I select two classifiers for my evaluation, SVM and Random Forest. Both of the classification models are trained and selected with leave-one-subject-out cross-validation. Moreover, I split 30% of the samples as my testing data for the performance evaluation.

4.3.2 DaLiAc Results

The overall performance (average accuracy classification rates of all activities) of SVM and RF using global (G) and combination of global and local features (G+L) for each sensor node is shown in Table 4.2. The average accuracy of SVM classifier using global features varies from 73% to 86% across different sensor positions. The best performance (86%) is achieved for 'Ankle' sensor, where 'Wrist' shows the worst (73%). Interestingly, leveraging local features with the global ones do not improve the performance of classifiers significantly for this classifier. In particular, the performance remains the same for 'Wrist' and 'Chest' sensors, where the performance rises by 2% and 1% for 'Hip' and 'Ankle' sensors, respectively. In comparison with SVM, the average performance of RF classifier varies from 78% to 89% across different sensor positions using global features. RF performs equal or even better to all sensors using the leveraged local and global features. The best improvement is found for 'Chest' and 'Hip' sensors, where the performance increases 5% by using RF classifier. Similar to SVM classifier, either the use of combination of global and local features improves the overall performance or they remain the same for individual sensor. The maximum improvement of performance (5%) for individual sensor using RF is higher than the one (2%) achieved using SVM. In addition, this maximum improvement using RF is found for 'Chest' and 'Hip' sensor, where for SVM it is found only for 'Hip' sensor.

In summary, leveraging global and local features can perform better or at least equal compared with the state of the art of using global features only in detecting uncountable activities. These results also indicate that the local features have

Table 4.2: Mean classification rates in DaliAc*

Sensor position	**Feature and classifiers**			
	SVM		**RF**	
	Global	**Global + Local**	**Global**	**Global + Local**
Wrist	73	73	78	78
Chest	84	84	81	85
Hip	81	83	81	85
Ankle	86	87	89	91

* Mean classification rates (in percent) of support vector machine (SVM) and Random Forrest (RF) classifiers using global features, and the combination of global and local features for individual sensor nodes wrist, chest, hip and ankle.

minimal capacity to extract additional information from the uncountable activities, which are exclusive in the DaLiAc dataset and thus the improvement of classification accuracy is limited. According to the definition of uncountable activities - they are difficult to be described by a set of gestures, which means there is no or very limited specific local patterns in them that can be extracted using my proposed local feature extraction methods.

4.3.3 AmA Results

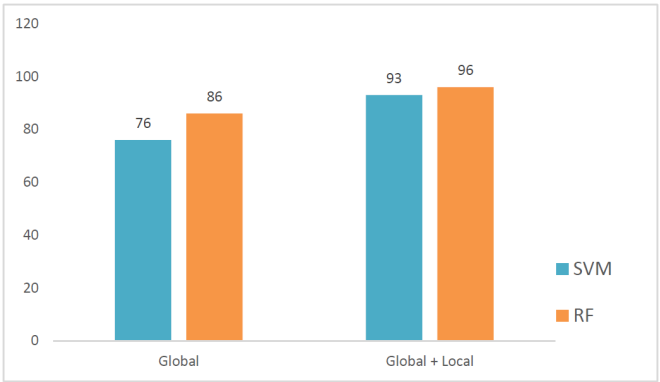
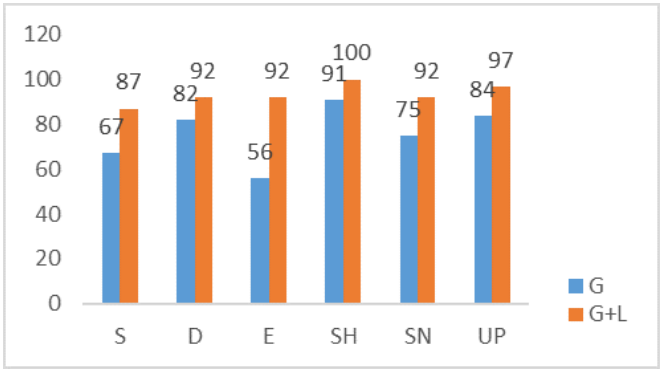


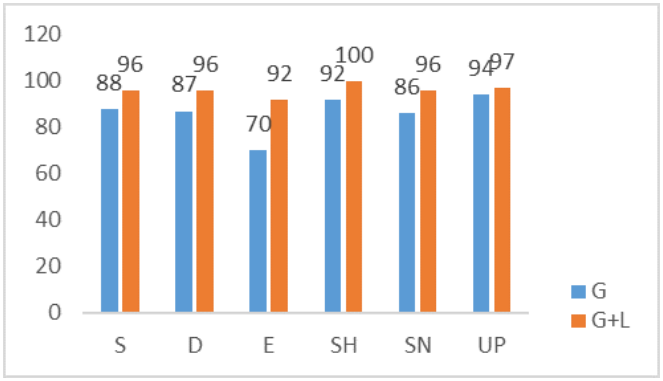
Fig. 4.7. Mean classification rates in AmA*

Legend Mean classification rates (in percent) of SVM and RF classifiers using global features, and the combination of global and local features.

The overall performance of SVM and RF classifiers using global (the state of the



(a) Performance of each activity in SVM



(b) Performance of each activity in Random Forest

Fig. 4.8. Performance of each Activity in AmA

Legend G: global features, G+L: the combination of the global and local features, S: smoking, D: drinking, E: eating, SH: scratching head, SN: scratching nose, UP: using phone

art) and combining global and local features are shown in Fig.4.7. Unlike DaLiAc dataset, the introduction of my proposed local features significantly improves the average classification performance of both classifiers for the AmA dataset, where the improvement is higher 17% for SVM (from 76% to 93%) compared to RF 10% (from 86% to 96%). For individual feature set (either global or leveraged), the RF classifier shows better performance than SVM, however the difference reduces to 3% from 10% when both local and global features are used. This indicates that the local features not only contribute to the improvement of performance but also improve the generalization capacity across machine learning models. This significant performance improvement can be attributed to the countable activities (described

by a specific number of gestures) of AmA dataset. Since local features extraction methods are proposed to capture temporal variation of an activity, they have added valuable information on top of global features. Therefore, the combination of global and local features provides a comprehensive description on the characteristics of an activity, which is translated in a significant improvement in classification performance.

As shown in Fig.4.8, from the individual activity point of view of AmA dataset, for both SVM and RF classifiers, the leveraged global and local features clearly show better performance than using global only. Surprisingly, the lowest classification performances of 56% and 70% are obtained for SVM and RF in classifying 'Eating' activity. Adding local features improves that to 92% for both classifiers, which is strikingly high and shows the strong capacity of local feature in improving performance for countable activities. All these results support my hypothesis that local features are important to be included with global features to better comprehend the characteristics of daily living activities, in particular of the countable activities.

In summary, the local features are more important for comprehending the characteristics of countable activity than the uncountable activities. In addition, the local features do not have any negative impact on classification performance when they are added with global features to classify uncountable activities. Thus, the use of local features on top of existing global features is always beneficial irrespective of the type of activities in a particular dataset. Most importantly, the true daily life activities are a combination of both uncountable and countable activities, hence combining global and local features to detect human life activities is necessary to obtain the optimal classification performance.

Chapter 5

MFE-HAR: Multiscale Feature Engineering for Human Activity Recognition Using Wearable Sensors

MFE-HAR approach [5][1] improves my previous work [4] to detect human daily activities using local and global features.

5.1 Multiscale Feature Engineering Design

5.1.1 Global Features

1. **Acceleration and Angular Velocity Based Features (AAVBF)**

   In previous studies, most of the features are statistical features which are extracted directly from accelerometer and gyroscope signals [4, 8]. In this work, I perform statistic analysis on acceleration to understand the time rate of change of velocity of arm movement while observing the rotation occurring from time t to $t+\Delta t$ by measuring the instantaneous angular velocity. Since velocity has a magnitude and a direction, acceleration and instantaneous angular velocity of arm movement can form two vectors $[x_{acc}, y_{acc}, z_{acc}]$, $[x_{gyo}, y_{gyo}, z_{gyo}]$ which contain both magnitude and direction. Therefore, features calculated from two vectors can be used to analyze the general movement

[1] This chapter is based on the under review paper [5]. The other co-authors are contributing to the writing of the paper while the supervisor is involved with the motivation, design and results discussion of the paper. I am responsible for the solution, implementation and experimentation.

trend (direction) of both acceleration and instantaneous angular velocity of arm movement. In AAVBF, I estimate arm movement from three-dimensional coordinate axes and their magnitudes. I extract my features by computing general statistics features (GSFs) including mean, standard deviation, maximum, minimum, skewness and kurtosis. Thus, my feature set consists of comprise 6×4 (3-D coordinate axes and magnitude) × 2 (accelerometer and gyroscope) = 48 features in acceleration and angular velocity based features.

2. **Position and Orientation Based Features (POBF)**

   Tondu et al. [21] indicates that normal person's arm movement can be accurately described as a seven-degrees-of-freedom pose movement which includes ball-and-socket joint, shoulder, upper arm, elbow, lower arm, wrist and hand. These seven parts collectively, which people named human pose, define the position and orientation of arm movement. Therefore, I can analyze the human pose to study arm movement. Moreover, study [22] reports the body poses based movement representations are highly effective for HAR. As a result, I use human pose for human behaviour analysis and calculate the position and orientation of arm movement through the data collected from Inertial Measurement Units, i.e., the accelerometer and gyroscope.

   Position Based Features: I estimate the integral of sensor reading using trapezoidal rule where an approximate area of the trapezoid in interval $[a_{i-1}, a_i]$. The velocity is calculated with an initial value of zero, given the vector of accelerometer reading and sensor's sampling time. After that, I obtain my position features with an initial value of zero, given the velocity and sampling time.

Orientation Based Features: Besides position features, I also extract orientation features. One common approach to represent orientation is Euler angles (pitch-yaw-roll) as shown in Fig.5.1. Pitch-yaw-roll is good for decomposing rotations into individual degrees of freedom. However, Euler angles are limited by a phenomenon called gimbal lock which prevents them from measuring orientation when pitch angle approaches ±90 degrees. Quaternion

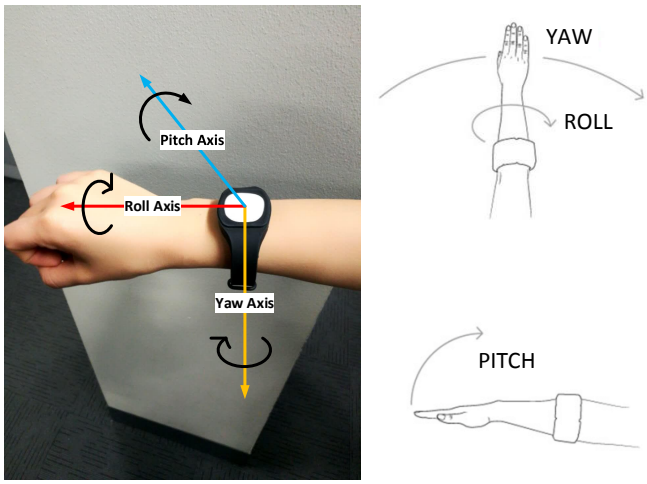


Fig. 5.1. Pitch-yaw-roll.

is an alternative approach for representing orientation. It is widely used in the development of unmanned aerial vehicles (UAVs) where people obtain a good feedback for the orientation controller [23]. In this study, quaternion is used to track the orientation of arm movement among various of human daily activities for high accurate rate instead of controlling orientation. The reason is that it avoids gimbal lock and is an efficient way to compute. After calculating quaternions in spatial rotations, I convert them to Euler angles directly to avoid gimbal lock. Quaternion is generally represented in the form:

$$q = (q_1, q_2, q_3, q_4) = (\cos\tfrac{\theta}{2}, \sin\tfrac{\theta}{2}nx, \sin\tfrac{\theta}{2}ny, \sin\tfrac{\theta}{2}nz) \tag{5.1}$$

I use first order Runge-Kutta to update q_0, q_1, q_2, q_3. The quaternion differential equation is listed below:

$$\begin{aligned}\begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix}_{t+\triangle t} &= \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix}_{t} \\ &+ \frac{\triangle t}{2} \begin{pmatrix} -gyro_x \cdot q_1 & -gyro_y \cdot q_2 & -gyro_z \cdot q_3 \\ +gyro_x \cdot q_0 & -gyro_y \cdot q_3 & +gyro_z \cdot q_2 \\ +gyro_x \cdot q_3 & +gyro_y \cdot q_0 & -gyro_z \cdot q_1 \\ -gyro_x \cdot q_2 & +gyro_y \cdot q_1 & +gyro_z \cdot q_0 \end{pmatrix}\end{aligned} \tag{5.2}$$

where $gyro$ is the angular velocity output of gyroscope. However, I do not use the gyroscope readings directly to calculate the quaternion in practice as they drift with time and generate additional errors accumulating over a period of time. In order to minimize the error, I use accelerometer and PI controller to calibrate the gyroscope so as to obtain an accurate angular velocity as shown in Fig.5.2, where initial values of $q0$, $q1$, $q2$ and $q3$ are 1, 0, 0, 0 respectively.

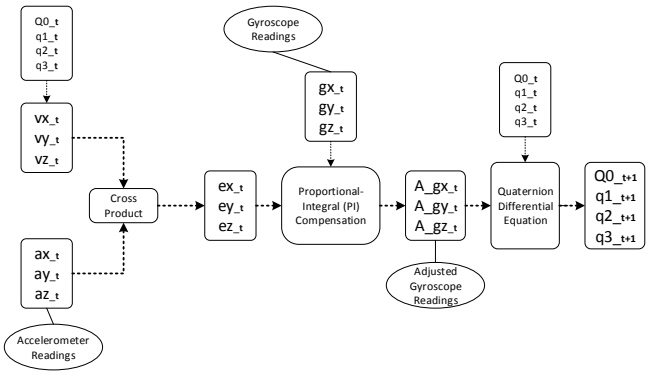


Fig. 5.2. Quaternion computation with PI cpntroller.

vx, vy and vz are calculated components of gravity, which can be computed by:

$$\begin{aligned} vx_t &= 2 \cdot (q1_t \cdot q3_t - q0_t \cdot q2_t), \\ vy_t &= 2 \cdot (q0_t \cdot q1_t + q2_t \cdot q3_t), \\ vz_t &= q0_t \cdot q0_t - q1_t \cdot q1_t - q2_t \cdot q2_t + q3_t \cdot q3_t. \end{aligned} \tag{5.3}$$

I calculate the errors between estimated components of gravity and calculated components of gravity by:

$$\begin{aligned} ex_t &= (ay_t \cdot vz_t - az_t \cdot vy_t), \\ ey_t &= (az_t \cdot vx_t - ax_t \cdot vz_t), \\ ez_t &= (ax_t \cdot vy_t - ay_t \cdot vx_t). \end{aligned} \tag{5.4}$$

I obtain adjusted gyroscope measurements by using noise filter with proportional-integral compensation to correct bias:

$$\begin{aligned} A_g x_t &= gx_t + k_p \cdot ex_t + exInt_t, \\ A_g y_t &= gy_t + k_p \cdot ey_t + eyInt_t, \\ A_g z_t &= gz_t + k_p \cdot ez_t + ezInt_t, \end{aligned} \tag{5.5}$$

and $exInt$, $eyInt$, $ezInt$ are integral error scaled integral gains:

$$\begin{aligned} exInt_t &= exInt_t + ex_t \cdot k_i, \\ eyInt_t &= eyInt_t + ey_t \cdot k_i, \\ ezInt_t &= ezInt_t + ez_t \cdot k_i. \end{aligned} \tag{5.6}$$

The initial values of $exInt$, $eyInt$ and $ezInt$ are 0, 0, 0, respectively. k_p is the proportional coefficient and k_i is the integral coefficient. In this study, I let K_p = 3.5 and $k_i = 0.05$ because these two values achieve the best performance comparing with other settings. Eventually, I use the adjusted gyroscope measurements to calculate the quaternions by (5.2). Then I use obtained quaternions to extract POBF. Since the exponential map for quaternion q is:

$$\exp(q) = \exp(0, \hat{n}\tfrac{\theta}{2}) = (\cos\tfrac{\theta}{2}, \hat{n}\sin\tfrac{\theta}{2}). \tag{5.7}$$

I calculate the logarithm of quaternion, which is:

$$\log(\exp(0, \hat{n}\tfrac{\theta}{2})) = (0, \hat{n}\tfrac{\theta}{2}), \tag{5.8}$$

where $\hat{n}\frac{\theta}{2}$ is the vector of the half value of the rotation angles in a quaternion. I calculate GSFs on $nx\frac{\theta}{2}$, $ny\frac{\theta}{2}$ and $nz\frac{\theta}{2}$, respectively, as part of orientation

features. In addition, I convert the quaternion to Euler angles by:

$$\begin{bmatrix} \varphi \\ \theta \\ \psi \end{bmatrix} = \begin{bmatrix} \arctan \frac{2(q_2q_3+q_0q_1)}{1-2(q_1^2+q_2^2)} \\ \arcsin[-2(q_1q_3+q_0q_2)] \\ \arctan \frac{2(q_1q_2+q_0q_3)}{1-2(q_2^2+q_3^2)} \end{bmatrix}. \tag{5.9}$$

I calculate GSFs on pitch, roll and yaw respectively as another part of orientation features. Eventually, I extract 1(position) + 6(GSFs) × 3($nx\frac{\theta}{2}$, $ny\frac{\theta}{2}$ and $nz\frac{\theta}{2}$) + 6(GSFs) × 3(pitch, roll and yaw) = 37 position and orientation based features.

5.1.2 Local Features

In addition to the global features, I also design local features, i.e., gesture features. Global features contain overall information of an activity while local features are more relevant to some specific points of an activity. In order to extract gesture features, I have to firstly identify the gestures in each activity. I make a regression data analysis on sensor readings to determine the gesture frames in an activity using Piecewise Linear Approximation techniques [24] as shown in Fig.5.3, which create the segmented version of the time series data, i.e., raw sensor data. In this work, I obtain the gestures from an activity using the bottom-up algorithm. I firstly create $\frac{n}{2}$ segments joining adjacent points in a n-length time series from each 5-second sliding window, and then calculate the approximation error of each initial segment. After that, the cost of merging adjacent segments is calculated for the purpose of merging the lowest cost pair of segments to a new and bigger segment. The new approximation error of merged segments are then updated. If the error is less than the defined stopping criteria, the process goes back to the first step and repeat the whole steps again, otherwise I obtain my gesture segments. In this process, time series data of segmented version are an approximation of the original series, which represent a liner regression of movement. I use the start and end anchors of

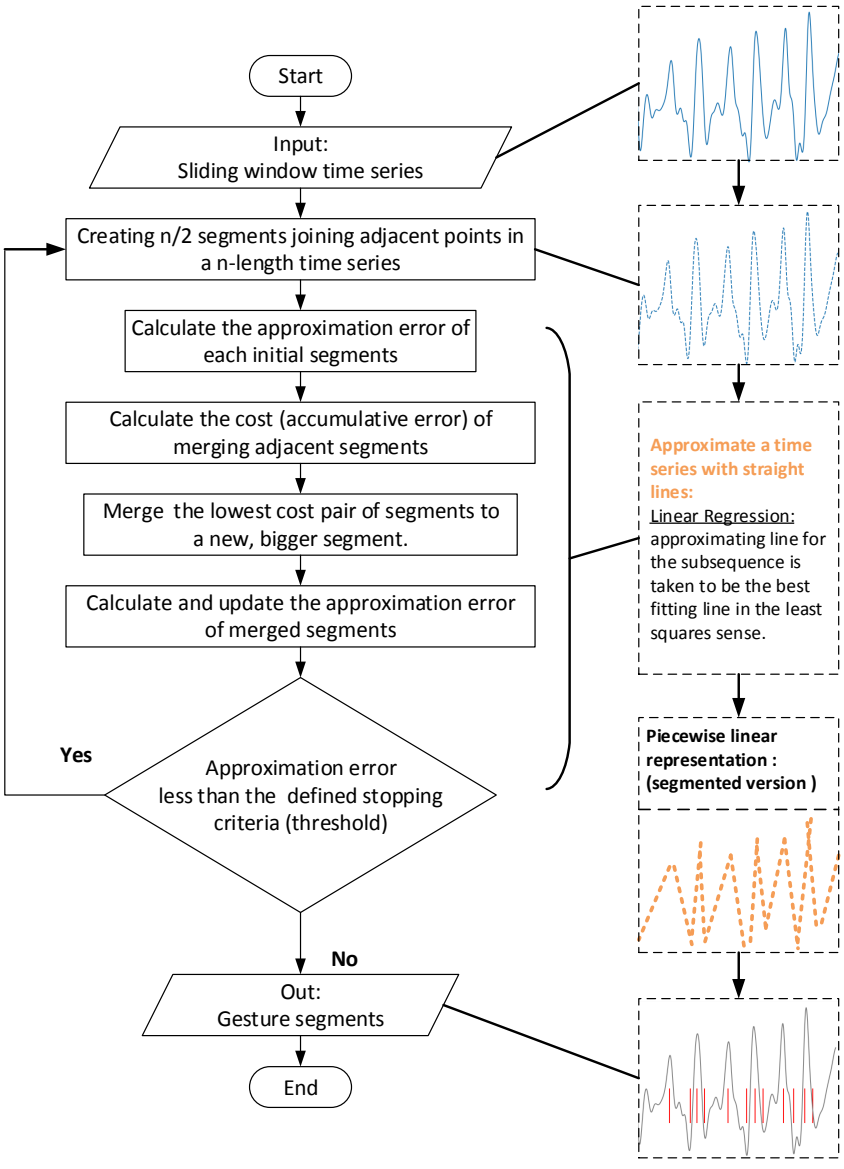


Fig. 5.3. Flowchart of Gestures Extraction

segmented version time series to separate signals to obtain the gesture frame. As an approximation can smooth the original series and reduce unwanted noise which help to calibrate accelerator's bias to a certain extent [25], it is unnecessary to apply the noise filter here. In this work, I obtain the gestures based on the change of the acceleration magnitude and direction to capture the general arm movement trend.

In each gesture frame, I convert the accelerometer and gyroscope values to Euler angles, and calculate their basic statistics including mean, maximum and minimum of roll, pitch and yaw, respectively. I use mean to analyze the general orientation of a gesture in a given time period while applying maximum and minimum to estimate the possible start and end of gesture orientation. Since the number of gesture segments is a variable across various activities or even for the same activity, the feature vector extracted from gestures cannot be used directly for training a classification model. I address this problem by calculating the statistic values' extrema, i.e., minima and maxima of the gesture segments for each activity. Since extrema can be used to compress a time series without losing too much important information [26], I use extrema of the gesture statistics to represent key gestures of an activity. Eventually, I extract 2 (minima and maxima) $\times$ (3 (max pitch, max yaw and max roll) + 3 (min pitch, min yaw and min roll) + 3 (mean pitch, mean yaw and mean roll)) = 18 local features in total.

5.2 Empirical Setting

In this study, my proposed MFE-HAR approach and baseline are verified using two datasets: Daily Life Activities (DaLiAc) [27] and mHealth [28]. The programming environment is Python 3.6, and the integrated development environment is Pycharm 2017. The computing configuration environment is Intel® Core™ i7-6700 CPU @ 3.40GHz processor, 16GB RAM and Windows 8.1 64-bit operating system.

5.2.1 Daily Life Activities (DaLiAc) Dataset

For more details of DaLiAc, see Section 4.2.1.

5.2.2 mHealth Dataset

This dataset includes 12 physical activities: standing still, sitting and relaxing, lying down, walking, climbing stairs, waist bends forward, frontal elevation of arms, knees bending (crouching), cycling, jogging, running, and jump front and back. The data are recorded by wearable devices placed on the chest, right wrist and left ankle of the subjects, where there are 10 participants. Multiple devices are able to measure the acceleration, the rate of run and the magnetic field orientation. In addition, 2-lead ECG measurements are provided by the device placed on the chest. For all sensing activities, the sampling rate is 50Hz, window length and overlapping are 4s respectively.

5.3 Results and Discussion

5.3.1 DaLiAc Dataset

In this section, I report arm-based activity recognition results by using five machine learning models: gradient boosting decision tree (GBDT), random forest (RF), logistic regression (LR), k-nearest neighbors algorithm (KNN) and support vector machine (SVM). My previous approach mentioned in chapter 4 is used to be my baseline. Fig.5.4 summaries their classification performance in terms of precision, recall and F-measure, based on my proposed MFE-HAR approach and baseline. It can be seen that the classification rates of MFE-HAR obtained from these models are all greater than 80%, where GBDT achieves the highest classification rate up to 93% in all accuracy measurements. Therefore, I select GBDT to compare the performance with the state of the art approaches [7, 8, 29, 30, 31, 32], where four wearable devices are placed on the right hip, chest, right wrist and left ankle.

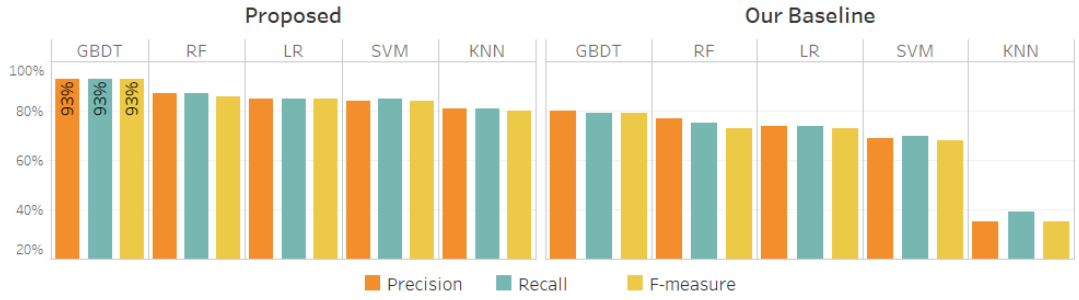


Fig. 5.4. Performance results of DaLiAc dataset.

Fig.5.5 shows the mean and each classification rate for 13 activities of my proposed solution compared to the state of the art. My proposed solution reaches the highest overall mean classification rate of 93%, and 100% in terms of RJ particularly, which outperforms other solutions in BC100, VC, BC50 and SW. The classification rates of AS, DS, ST, SI, WD, RJ, LY, WK and RU are very close to [8], which is only lower by 2% on average. However, my MFE-HAR approach is superior to previous studies, such as [8] which uses multi-devices, while my proposed solution is based on a single-device on wrist to detect activities.

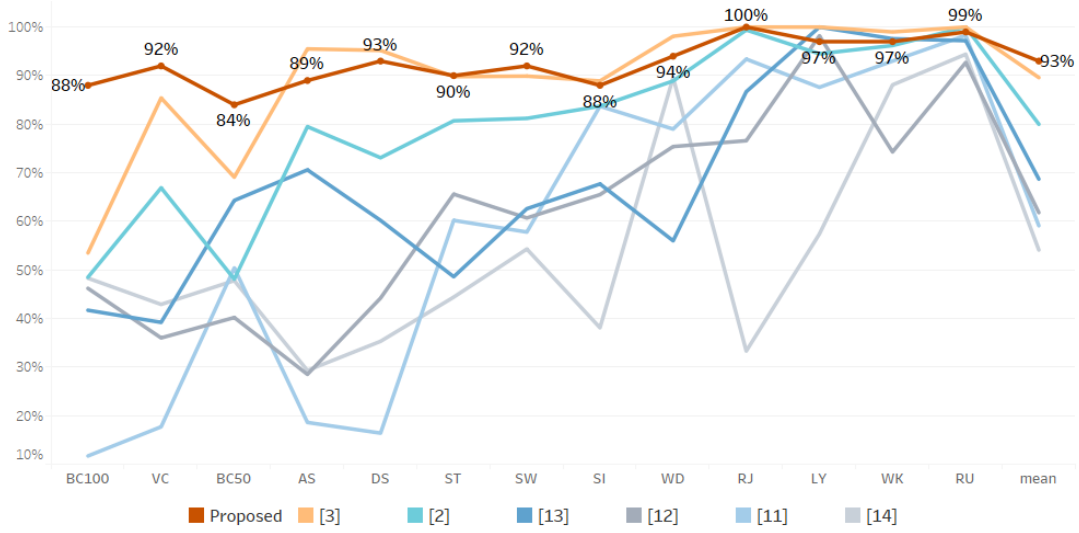


Fig. 5.5. Classification rates comparison of DaLiAc dataset.

Study [8] has a better performance on LY, WD, SW, WK, AS and DS due to multi-

devices are able to provide comprehensive information to represent human activities. One analysis [33] based on the performance of each single device deployed in multi-device solution indicates that in DaLiAc most considered activities are strongly related to the movements of lower extremities, while the hip covers a wide range of basic activities. That is, ankle and hip are two essential positions to recognize human activities. Therefore, the devices placed on ankle and hip provide more significant contributions to HAR in DaLiAc, comparing with devices placed on wrist. My proposed MFE-HAR approach can achieve similar or even better recognition results than multi-device based solutions using a single device placed on wrist. The reason is that most studies only use AAVBF to recognize human activities which can be only used to analyze the characteristics of acceleration and angular velocity of an activity. These features cannot be used to determine the movement state in a 3D space. I observe that tracking arm movement in a 3D space can be actually treated as tracking a wearable device's movement. If I treat a wearable device as a rigid body, I would need position and orientation to determine the movement state in 3D space [34]. As a result, not only acceleration and angular velocity, but position and orientation are used in my proposed MFE-HAR solution to capture arm movement.

For bicycling on ergometer (50W) and bicycling on ergometer (100W), I make a significant improvement on their recognition accuracy comparing with the state of the art solutions. According to confusion matrix of my MFE-HAR classification results as shown in Fig.5.6, I find that most instances of BC50 and BC100 are misclassified to each other. There are 84% correctly classified BC50 with lower resistance level and 11% misclassified as BC100 with higher resistance level. Meanwhile, 88% BC100 are correctly classified and 8% misclassified as BC50. The reason is that they are actually confounding activities with different resistance levels but similar movement patterns to confuse the classifier. I analyze top 20 important features in recognizing the activities between them using Extra Trees classifier and find out that 12 features are local features, which indicates that MFE-HAR has an outstanding

ability to recognize confounding activities by combining local features with global features.

	SI	LY	ST	WD	VC	SW	WK	AS	DS	RU	BC50	BC100	RJ
SI	0.88	0.04	0.00	0.01	0.00	0.03	0.00	0.00	0.00	0.00	0.01	0.03	0.00
LY	0.03	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ST	0.02	0.00	0.90	0.01	0.00	0.03	0.00	0.00	0.00	0.00	0.01	0.03	0.00
WD	0.00	0.00	0.01	0.94	0.01	0.02	0.00	0.00	0.00	0.00	0.02	0.01	0.00
VC	0.00	0.00	0.00	0.00	0.92	0.07	0.01	0.00	0.00	0.00	0.00	0.00	0.00
SW	0.00	0.00	0.01	0.01	0.05	0.92	0.00	0.01	0.00	0.00	0.00	0.01	0.00
WK	0.00	0.00	0.00	0.00	0.00	0.01	0.97	0.01	0.00	0.00	0.00	0.00	0.00
AS	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.89	0.00	0.00	0.00	0.00	0.00
DS	0.00	0.00	0.00	0.00	0.00	0.02	0.05	0.00	0.93	0.00	0.00	0.00	0.00
RU	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.99	0.00	0.00	0.00
BC50	0.00	0.00	0.00	0.01	0.00	0.02	0.02	0.00	0.00	0.00	0.84	0.11	0.00
BC100	0.00	0.00	0.01	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.08	0.88	0.00
RJ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Fig. 5.6. Confusion matrix of MFE-HAR classification accuracy for DaLiAc.

5.3.2 mHealth Dataset

I also use mHealth dataset to estimate the validation of MFE-HAR approach. The recognition accuracy results are shown in Fig.5.7. It clearly indicates that by using even only one wrist device, my proposed solution performs better than multi-device based approach [35] and my baseline on all five machine learning models. The GBDT model achieves the highest mean classification rate 98% compared with other four models, where recognition performance is good for confounding activities such as walking and climbing stairs, jogging and running. Therefore, my MFE-HAR approach is able to be applied in different wearable devices.

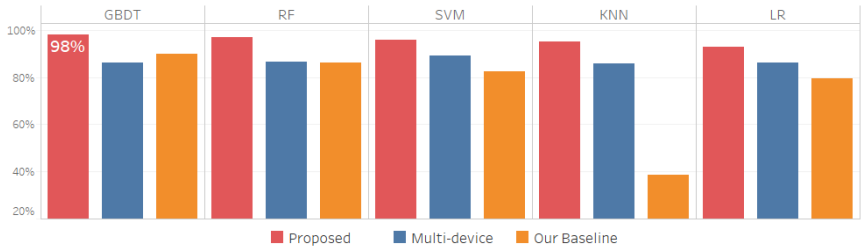


Fig. 5.7. Accuracy results comparison of mHealth dataset.

Chapter 6

Conclusion

In spite of many existing works on human activity recognition, few solutions are practical enough to be adopted in mission critical real-world applications. Especially in the health domain, these solutions either require excessive number of wearable devices (multi-device approach) or demand complex classification models and large training datasets (deep classification approach). My studies attempt to look into the very nature of daily activities and propose effective and efficient feature engineering approaches to address the fundamental challenges in HAR. My experiments have demonstrated the proposed methods are superior to the state of the art. As a next stage work, I intend to incorporate the feature engineering approaches into a real-world IoT health project to investigate its applicability with more complex scenarios at real time.

Bibliography

[1] Babin Bhandari, JianChao Lu, Xi Zheng, Sutharshan Rajasegarar, and Chandan Karmakar. Non-invasive sensor based automated smoking activity detection. In *Proc. of EMBC*, pages 845–848. IEEE, 2017.

[2] Rongjunchen Zhang, Xiao Chen, Jianchao Lu, Sheng Wen, Surya Nepal, and Yang Xiang. Using ai to hack ia: A new stealthy spyware against voice assistance functions in smart phones. *arXiv preprint arXiv:1805.06187*, 2018.

[3] Jianchao Lu, Jiaxing Wang, Xi Zheng, Chandan Karmakar, and Sutharshan Rajasegarar. Detection of smoking events from confounding activities of daily living. In *Proc. of ACSW*, page 39. ACM, 2019.

[4] Jianchao Lu, Xi Zheng, Jiaxing Wang, and Jiong Jin. Leveraging local and global features to detect human daily activities. *ACM Transactions on Internet of Things*, 2019, Under Review.

[5] Jianchao Lu, Xi Zheng, Jiaxing Wang, Wanlei Zhou, and Michael Sheng. MFE-HAR: Multiscale feature engineeringfor human activity recognition using wearable sensors. In *Proc. of MobiQuitous*. ACM, 2019, Under Review.

[6] J Hayward, G Chansin, and H Zervos. Wearable technology 2017–2027: Markets, players, forecasts. *IDTexEx Report*, 2017.

[7] Ling Bao and Stephen S Intille. Activity recognition from user-annotated acceleration data. In *Proc. of PerCom*, pages 1–17. Springer, 2004.

[8] Heike Leutheuser, Dominik Schuldhaus, and Bjoern M Eskofier. Hierarchical,

multi-sensor based classification of daily life activities: comparison with state-of-the-art algorithms using a benchmark dataset. *PloS one*, 8(10):e75196, 2013.

[9] Nils Y Hammerla, Shane Halloran, and Thomas Plötz. Deep, convolutional, and recurrent models for human activity recognition using wearables. *arXiv preprint arXiv:1604.08880*, 2016.

[10] Jason Brownlee. Discover feature engineering, how to engineer features and how to get good at it. *Machine Learning Process*, 2014.

[11] Sumeyye Konak, Fulya Turan, Muhammad Shoaib, and Özlem Durmaz Incel. Feature engineering for activity recognition from wrist-worn motion sensors. In *Proc. of PECCS*, pages 76–84, 2016.

[12] Qu Tang, Damon J Vidrine, Eric Crowder, and Stephen S Intille. Automated detection of puffing and smoking with wrist accelerometers. In *Proc. of PervasiveHealth*, pages 80–87. ICST, 2014.

[13] Yu-Liang Hsu, Shih-Chin Yang, Hsing-Cheng Chang, and Hung-Che Lai. Human daily and sport activity recognition using a wearable inertial sensor network. *IEEE Access*, 2018.

[14] Hiram Ponce, Pedro Ponce, Héctor Bastida, and Arturo Molina. A novel robust liquid level controller for coupled-tanks systems using artificial hydrocarbon networks. *Expert Systems with Applications*, 42(22):8858–8867, 2015.

[15] Eamonn Keogh, Selina Chu, David Hart, and Michael Pazzani. An online algorithm for segmenting time series. In *Proc. of ICDM*, pages 289–296. IEEE, 2001.

[16] Eamonn Keogh, Selina Chu, David Hart, and Michael Pazzani. Segmenting time series: A survey and novel approach. *Data mining in time series databases*, pages 1–21, 2004.

[17] Johan Himberg, Kalle Korpiaho, Heikki Mannila, Johanna Tikanmaki, and Hannu TT Toivonen. Time series segmentation for context recognition in mobile devices. In *Proc. of ICDM*, pages 203–210. IEEE, 2001.

[18] Stan Salvador and Philip Chan. Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5):561–580, 2007.

[19] Mlxtend (machine learning extensions). `https://github.com/rasbt/mlxtend`, 2017.

[20] scikit learn. `http://scikit-learn.org/stable/`, 2017.

[21] Bertrand Tondu, Serge Ippolito, Jérémie Guiochet, and Alain Daidie. A seven-degrees-of-freedom robot-arm driven by pneumatic artificial muscles for humanoid robots. *The International Journal of Robotics Research*, 24(4):257–274, 2005.

[22] Cheng Chen, Alexandre Heili, and Jean-Marc Odobez. A joint estimation of head and body orientation cues in surveillance video. In *Proc. of ICCV Workshops*, pages 860–867, 2011.

[23] Redouane Dargham, Adil Sayouti, and Hicham Medromi. Euler and quaternion parameterization in vtol uav dynamics with test model efficiency. *coordinates*, 2(3):4, 2015.

[24] Tom Schoellhammer, Ben Greenstein, Eric Osterweil, Michael Wimbrow, and Deborah Estrin. Lightweight temporal compression of microclimate datasets. 2004.

[25] Om Prakash Yadav and Shashwati Ray. Smoothening and segmentation of ecg signals using total variation denoising–minimization-majorization and bottom-up approach. *Procedia Computer Science*, 85:483–489, 2016.

[26] Eugene Fink and Harith Suman Gandhi. Important extrema of time series. 2007.

[27] DaLiAc-Daily Life Activities. `https://www.mad.tf.fau.de/research/activitynet/daliac-daily-life-activities/`, 2013.

[28] MHEALTH Dataset Data Set. `http://archive.ics.uci.edu/ml/datasets/mhealth+dataset`, 2014.

[29] Nishkam Ravi, Nikhil Dandekar, Preetham Mysore, and Michael L Littman. Activity recognition from accelerometer data. In *Proc. of AAAI*, volume 5, pages 1541–1546, 2005.

[30] Juha Parkka, Miikka Ermes, Panu Korpipaa, Jani Mantyjarvi, Johannes Peltola, and Ilkka Korhonen. Activity classification using realistic data from wearable sensors. *IEEE Transactions on information technology in biomedicine*, 10(1):119–128, 2006.

[31] Stephen J Preece, John Yannis Goulermas, Laurence PJ Kenney, and David Howard. A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Transactions on Biomedical Engineering*, 56(3):871–879, 2009.

[32] Shaopeng Liu, Robert X Gao, Dinesh John, John W Staudenmayer, and Patty S Freedson. Multisensor data fusion for physical activity assessment. *IEEE Transactions on Biomedical Engineering*, 59(3):687–696, 2012.

[33] Dominik Schuldhaus, Heike Leutheuser, and Bjoern M Eskofier. Towards big data for activity recognition: a novel database fusion strategy. In *Proc. of BODYNETS*, pages 97–103. ICST, 2014.

[34] David Baraff. Physically based modeling: Rigid body simulation. *ACM SIGGRAPH Course Notes*, 2(1):2–1, 2001.

[35] Eftim Zdravevski, Petre Lameski, Vladimir Trajkovik, Andrea Kulakov, Ivan Chorbev, Rossitza Goleva, Nuno Pombo, and Nuno Garcia. Improving activity recognition accuracy in ambient-assisted living systems by automated feature engineering. *IEEE Access*, 5:5262–5280, 2017.