Ingenieurpraxis
realtime FoG detection and hardware implementation

Wei Zhaoxiong
ge25zir@mytum.de
Technische Universität München

September 7, 2020

Abstract

Freezing of gait (FoG) is an abnormal gait pattern that accompany Parkinson's disease(PD). During such short, temporary episodes, the patient is not able to move the feet forward despite the intention of walk. According to research [1], a sequence of rhythmic stimuli is able to help patient to walk again during FoG. Therefore, it is important to detect FoG pattern at the very beginning or even before it comes.

In this paper, we tried to extract and select gait features using DAPHNET dataset. A fisher discrimination analysis and feature thresholding is applied to further reduce its dimension. Besides, we proposed a new evaluation standard that focus on improving the user experience and tune relevant parameters based on it. In the end, the algorithm is implemented in a source-constrained device.

Keywords – Freezing of gait, feature extraction, feature selection, gait features, fisher discrimination analysis, DAPHNET

Introduction

Parkinson's disease (PD) is the second most common aging-associated neurodegenerative disorder [2]. It is characterized by motor features such as muscular stiffness, resting tremor, hastening of the gait(festination) and poor postural stability. Freezing of gait (FoG) is one of the symptoms of PD, during such sudden and short episode, the patient experiences gait disturbance ranging from complete sudden akinesia to milder leg trembling or short shuffling steps events, usually described by patients as feeling the feet stuck to the floor. These symptoms result in emotional stresses and reduced quality of life [3]. However,there is still no effective treatment or drug-resistant against such disease [4].

The work of Hausdorff et al. shows rhythmic auditory stimulation (RAS) was shown to be particularly effective at improving gait among PD patients [5]. In their experiment, regular metronome ticking sounds were applied as RAS with a rate of 110% compared to the natural walking rate of tested patient.This served to improve gait stability and speed.

Since unnecessary ticking sounds during normal walking can be annoying for patient and we want to only provide RAS during an actual or impending FoG event, a realtime detection of FoG is important.

A variety of methods have been proposed to detect FoG. Moore et al. used an Inertial Measurement Unit (IMU) mounted around a shank, and defined a freeze index(FI) using power spectrum analysis of accelerometer signals [6]. Each patient use individually calibrated thresholds to recognize FoG. The experiment shows that 89% of FoG event can be detected, However, 10% of non-FoG events can generate false alarm.

Based on this, Bachlin et al proposed a real-time FoG detection algorithm combines the FI with a second power threshold to discriminate FoG from normal gait, resulting in higher sensitivity (88.6%) and specificity (92.4%) [7].

Besides, there are also several machine learning based approaches that exploiting the data set. Florenc et al. applied kernel-lda to the accelerometers' raw data to reduce the dimension of features. Then a k-nearest neighbor algorithm (k-NN) was used to classify gait in pre-FoG, no-FoG and FoG [8].

Natasa et al. extract several features from raw data and using Boruta algorithm to reduce the dimension, they make a comparison of result using classifiers like Support Vector Machine, Random Forest and k-NN [9]. Similarly, Parisa et al. collected inertia and physiological signals to extract freezing patterns, the set of features are fed into a Fully-Connected-Neural Network(FCNN) to learn and predict freezing episodes [10].

Considering that we are using source-constrained wearable devices, some methods can be too expensive to implement, such as parameters of FCNN takes too much memory space and computational resources and it may hard to meet the requirement of real-time. We are not going to consider it as an option in our setup.

Approaches Pursued

In this research, I tried several methods mentioned before, for some paper I cannot get the desired result because of insufficient understanding of background. In the end, I find a method which yields the best result and I'm going to focus more on talking about it.

The DAPHNET Freezing of Gait dataset is used for this research. This dataset consists of sensor data extracted from individuals with PD. Three tri-axial accelerometers were placed at the shank (just above the ankle), thigh (just above

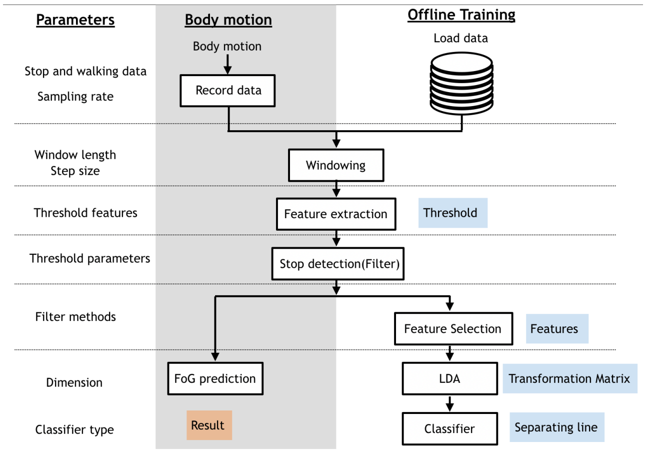


Figure 1: Working flow

knee) and lower back to collect acceleration in three axes with a sample rate of 64Hz.

The dataset include data from 10 patients, they are asked to perform walking in straight line, walking with numerous turns and some daily living tasks. The label is done by two physiotherapists using video recording. During the test, two patient did not show any symptoms of FoG, so they are excluded in this research. In total this dataset collected five hundred minutes of data and 237 FoG events happened within this time, half of the FoG episodes lasted less than 5s and most part lasted less than 20s (93.2%). Therefore, this is an unbalanced dataset who has much more non-FoG periods than FoG episodes.

In the first part of this research internship, I tried to implement the algorithm of Florenc. However, I failed to separate classes in the test set by using kernel-lda. Also, the author informs that his work does not perform well when it comes to larger data set. So I stopped on working with this algorithms.

Another machine learning algorithm I tried is Neural Network(NN), however, with limited number of weights and layers allowed, it's hard to do the classification correctly.

According to my observation, the reason for the failure of employing machine learning methods could be the followings. First of all, the data set is not quite balanced, the FoG labeled data is far less than non-FoG data set. Second, there might be some miss-labeled data. Since the data set is labeled with eye observation, it is possible the detection of FoG is slower than it should have been for second. Considering that machine learning methods are purely data-driven, such problems with the dataset could have huge impact on the result.

After a conversation with my advisor, I change the research target to gait cycle analysis. This is a study that is used to assess and treat individuals with conditions affecting their ability to walk. From the data set, it is possible to extract features like walking speed and cadence. After some gait features are extracted, thresholds and machine learning techniques are applied. The general procedure of this method is shown in figure 1. More detail information can be found in next section.

Method

As mentioned before, some methods focusing on finding a good feature and used thresholds to detect FoG episodes, while the others applied machine learning methods to extract features and do classification. In the proposed method, we combine these two method together.

According to previous experience, it is hard to use LDA or PCA directly on raw data or low-pass filtered data to extract features with generality because of the unbalanced dataset and noises within. However, the result with test set improves a little when LDA is performed with extracted features. Therefore, as figure 1 shows, after a window pre-processing step, several features are extracted from it.

Another thing need to be done is data cleaning, just like the method proposed by Bachelin et al.[7], we use total energy in frequency band [0-8Hz] to detect standing period and filter it away. This makes the dataset less unbalanced and reduce computation load at the same time.

In the next step, we use feature selection methods to filter away features on different channels that are not efficient. Since the FoG feature do not express itself equally on all axis. Also, different patient has different symptoms during FoG.

After the feature dimension is reduced, LDA is applied to remained good features to reduce the dimension, followed by a decision tree for classification. It is also possible to apply SVM, random forest and other machine learning classification methods.

Data Pre-processing

The length of window and step length are variables in several researches, it is possible to do a grid search for a best fit. There's also a trade off between window size and reaction time. We do not put too much effort in these parameters and set them as fixed which performs well for several other methods.

The sliding window is chosen as 2 seconds with 256 * 9 floats per window. The step length is chosen as 0.5 seconds. They are going to be the basis for feature extraction in next step.

Another task in this step is labeling, as mentioned before, there is a huge unbalance between FoG labeled and non-FoG labeled data. Since we would like to detect the beginning and impending FoG episodes, as long as one sampling point is labeled as FoG, the whole window will be labeled as FoG. Besides, for each non-FoG labeled window before a FoG labeled window, we also label it as FoG window, this reduce the number of non-FoG window and increases FoG windows.

Feature Extraction

There are lots of researches focus on features of FoG episodes, the feature extraction can be performed in both time domain and frequency domain.

From the observation of acceleration data in time domain, we can visualize clear some periodic pattern during the non-

FoG phase, However, such characteristic is less obvious or not existing in FoG phase.The following three methods can be applied to exploit this characteristic.

1. Sample Entropy: A measure of complexity, Often used to analyze the physiological variability in human gait and it is derived from approaches developed by Richman and Moorman[11]. A feature which shows the repeatability and predictability within one window. The formula used to calculate it is:

$$\mathbf{SampEn(m, r, N)} = -log(A^m(r)/B^m(r)) \quad (1)$$

   In this study the dimension m is selected as 2, r as 0.2*δ, where δ is standard deviation.

2. Cross Correlation :Another method to exploit the periodicity is to use cross correlation. This shows the similarity between a signal and its shifted version. The window size in our case is selected as 2 seconds, so generally 2 to 3 steps can be observed in it. In this way,a shift with highest value can be considered as a step interval.

$$\mathbf{CrossCorrelation} = \sum_{n=0}^{n=len} f[n]g[n+\delta] \quad (2)$$

   Besides, the highest correlation value also shows the possibility of existing a periodic pattern in this window, which can be used as another feature.

3. Fast step detection: Both methods mentioned before is quite expensive in terms of time, and when it is implemented on hardware the realtime capability might not be guaranteed.

   Therefore, we invented a more light-weight method based on the observation from data. A step usually comes with a large drop and rise in the acceleration data as figure 2b shows, such U shape can be used as a feature to detect one step. As the figure shows, we can search for continuous decrease in the value and find potential step points. With such information we can find the step interval, step frequency. The threshold to get classified as a U shape is different individually and need to be calculated using normal walking gait data, such as the blue part in figure 2a.

   The result is as the figure 2c shows, between 50 and 250 we can clearly see the patient is walking normally, so the step interval most of time stays at 60ms. Whenever the patient turns, we can clearly observe some gap. When the patient experience FoG, the step interval can decrease to 0 for patient 2.

From another point of view, we can also extract useful information from frequency domain.

1. Dominant frequency peak: Dominant frequency is defined as the highest magnitude frequency in a power spectral density. This is the frequency which carries the most energy. Generally speaking, this value should coincide with step speed of walking during normal walking, which is 1-2 Hz.

Features in time domain	
Features	Description
sample entropy	Measures repeatability or predictability within one window :
step interval	Time interval between two steps in one window
max correlation	Measures the periodicity in one window
step depth	Difference between max and min value in one window
step counts	Counts of steps in one window
variogram	The measure of smoothness of data in time series
portion above mean	The proportion above the mean of the observationswithin the window whose values are greater than the mean of the window
Features in frequency domain	
Features	Description
loco-band energy	energy in [0.5-3Hz] frequency band
loco-band+freeze-band energy	energy in [0-8Hz] frequency band
freezing index	The power in the 3-8Hz band divided by the power in 0.5–3 Hz band
dominant frequency	The frequency with maximal Power Spectral Density (PSD)
wavelet mean	the mean of coefficient of DWT in the third level, which represents energy in freeze band

Table 1: Features

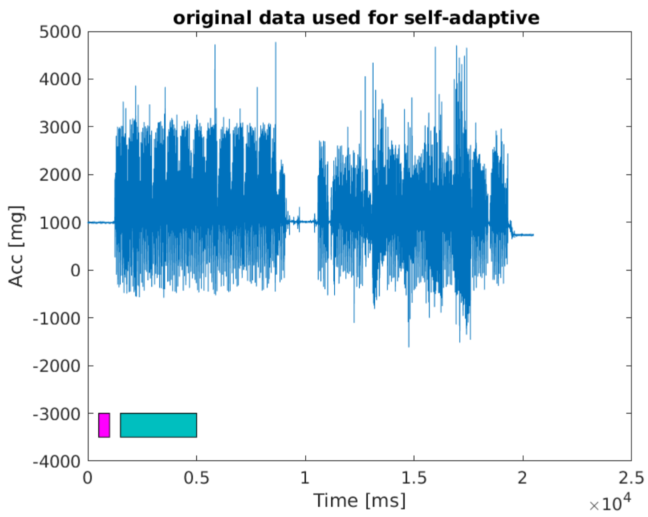
2. Freezing Index and energy thresholds: Moore et al. discovers that more energy components appears in [3-8Hz] band during FoG [6]. They introduced a freeze Index(FI) as a strong indicator of FoG.

   To calculate FI, we apply a method proposed by Baechelin et al., they used a NFFT-point FFT to get power spectral density [7], where NFFT = window length * sampling frequency. The locomotion energy is calculated as the integral in [0-3Hz], and freeze energy is integrated in [3-8Hz]. The freeze Index is calculated as :

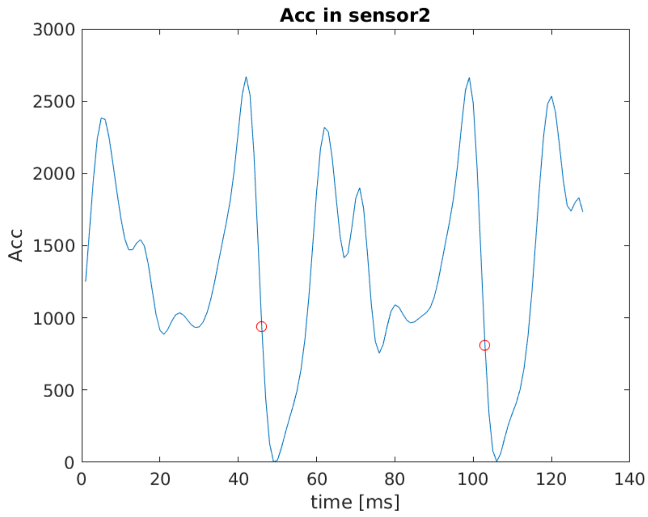
$$\mathbf{FI} = loco/(loco + freeze) \quad (3)$$

3. Wavelet mean: Wavelet transform is a standard tool that shows how the power amplitude of a specific frequency in a signal change over time. In the paper of Rezvanian and Lockhart et al. [12], they used the ratio of sum of wavelet coefficient in freeze and locomotion bands to discriminate FoG and non-FoG events. Here we used the mean of absolutes values of Discrete Wavelet Transform(DWT) from level three which includes freeze band. Daubechies wavelet of order four is chosen as mother wavelet.

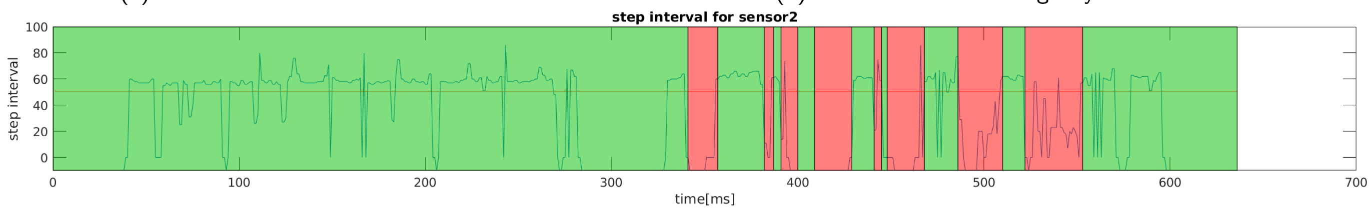
There are also some other features, a summary is found in table **??**.



(a) acceleration in x axis of ankle sensor



(b) acceleration of walking in y axis of ankle sensor



(c) measured step interval

Figure 2: The red area shows FoG, the green area shows non-FoG(a)The acceleration in y axis of ankle sensor, the purple part shows stop, the blue part shows normal walking.(b)Zoom into one of the walking window, two U shape can be observed, the red point shows the detected step.(c)The time interval between two detected points shows where the FoG might be

Thresholds and Feature Selection

To solve the unbalanced data problem, it is necessary to do cleaning to dataset before processing it. According to the measurement of Baechlin et al, when the patient is standing still, its energy between [0-8Hz] frequency band is significantly lower than other status. Therefore, we can first exploit such characteristic and filter out windows when patient is standing and only make classification between FoG and walking or turning. It is also possible to try other features as thresholds, however, we need to make sure that the filtered region is not broken into small pieces so that the time continuity around FoG is still valid, this guarantees an evaluation standard that we are going to talk about in the next part.

After the features are extracted, and standing windows are filtered away. generally we can just feed them into a dimension reduction tool like LDA. However, the result of using features directly does not generate good results.

According to the observation, different features have different dependencies with labeling. Also, different patients have different gait features during the FoG, therefore, it is important to find out which features are more efficient for each patient. This can be explained with the fact that different patients have different symptoms during the FoG, according to the research of Parisa Tahafchi et al.[10] FoG can be shown as festination, akinesia and trembling.

Although LDA is capable of selecting not meaningful features out by giving them very small weights, but this ability is fully depend on the quality of data set.

We selected filter method to do feature selection, because it selects features independently for any machine learning algorithm. It can help us remove features that contains little information.

Correlation filter method such as Spearman's rank correlation can be used to measure the degree of association between two set of variables with a monotonic function. It is capable of handling non-linear functions in compare with Pearson correlation.

Statistical method such as mutual information a measure of the mutual dependence of two variables. It measures the amount of information obtained about one variable through observing the other variable.

In the experiment we are going to use both of them and make comparison.

Feature reduction

After we get the cleaned dataset and selected features, we are ready for LDA. Since simple thresholds on all these features is still too complicate and computational expensive. We still need this step to further reduce the dimension down to a level which is easier for classification.

The output dimension of LDA is depends on classification method. Higher dimension comes with higher performance at the cost of more computation and memory resource requirement.

The simplest version of classifier is dealing with one dimensional features. In such case, the classifier only need to draw a line and discriminate the data based on this threshold, as in figure 3a. The last step is to design a classifier which gives us this line. It is worth noted that the design of classifier is closely coupled with evaluation standard, which need to be clarified first.

Classification and Evaluation standard

Most papers regarding the FoG detection uses classical classification quality metrics,such as precision, specificity Recall and F1-Score.

However, such classification standard does not consider the user experience and application characteristic. From our observation, there are some phases during FoG where the result can be different. The beginning of a FoG phase has different characteristic than the other part and special detector can be made to detect these part. Since we don't know how long this part is, we can only try to modify our detector to exploit such feature.

From one side, since the objective is to detect FoG in a real-time way, it is desired that the time window before the actual FoG is also classified as a FoG.

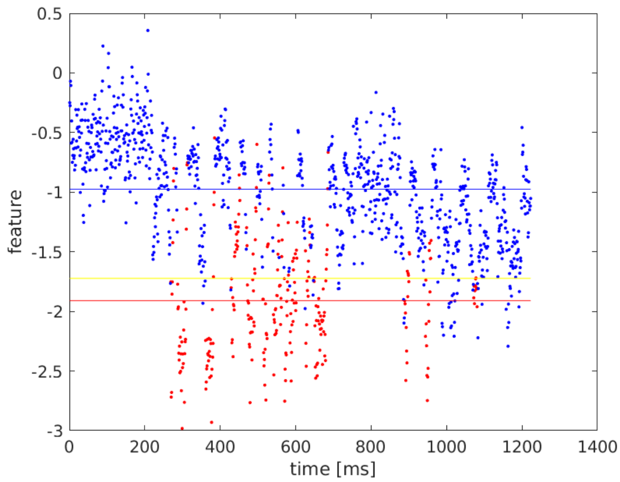
From the other side, the detector may label several window that are found inside a FoG labeled sequence as non-FoG, however, such false detected points does not have a high cost in real life, since the patient can recover from the FoG as long as he get an alarm for a certain time in the beginning part. It is more important not to make false alarm when the patient is walking normally.

Besides, considering that a false alarm happens just after the FoG is not so annoying as a false alarm during a normal walk, we also consider it as a correct label.

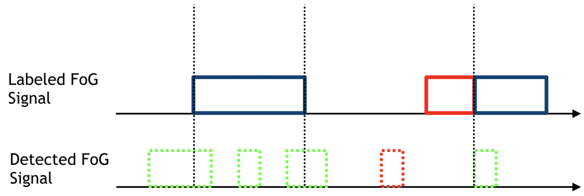
In conclusion, the following situations need to be reconsidered, as shown in figure 3b:

1. The FoG rising edge: If one continuous window sequence before the rising edge is classified as a FoG period, then the whole window sequence is considered as correct labeled even part of them are not labeled as FoG in the dataset.
2. During the FoG period: If the edge is detected and solved with one filter window, we consider that the patient would have stop experiencing the FoG period, therefore, the whole continuous FoG period is considered as solved. Otherwise, if the FoG is detected after the coming edge, only part of the FoG period is considered as solved
3. Whenever one FoG period ends, if one continuous window sequence is correct labeled at this point, then all filter window continuously followed are considered as correct.

In this way, such evaluation standard does not punishes the wrong results that after a correct FoG beginning detection and reward the part the detection before beginning of FoG. We make the detector focus more on the beginning part of FoG phase instead of whole part, It also makes less annoying false alarm. In this way, we have a new evaluation standard. New Precision stands for the percentage of correct alarm in all windows labeled as FoG, which is similar to Precision. New Recall stands for how many FoG windows are solved,



(a) lda feature reduced to one dimension



(b) Evaluation method

Figure 3: (a)The blue line shows the mean value of features in non-FoG windows, the red line shows the mean value of features in FoG windows. The target is to find a separating line(yellow) which best discriminate the data

which is the similar as recall before.

$$\textbf{New Precision} = \frac{\sum corrcect\,FoG\,alarms}{\sum all\,FoG\,alrams} \tag{4}$$

$$\textbf{New Recall} = \frac{\sum solved\,FoG}{\sum all\,FoGs\,exists} \tag{5}$$

$$\textbf{New F1} - \textbf{Score} = 2 * \frac{New\,Precision * New\,Recall}{New\,Precision + New\,Recall} \tag{6}$$

To find a separating line, we apply the same method as in decision tree. Several possible thresholds between the mean of FoG features and non-FoG features is selected, then the quality metrics is calculated to find out the best threshold.

In decision tree algorithm, generally gini index is going to be used, However, it does not performs so well with unbalanced data since the line is always too agressive with high recall and low precision. So we tried to use a scaled version of gini index in which we make each FoG points n times bigger in terms of number, n is the ratio of number between the non-FoG points and FoG points. Also we can use the new evaluation standard and try to find a best result with highest F1-score.

Result

The first result worth mentioned is feature selection. Using statistical tools we can figure out which features on which channel is really closely correlated with FoG symptoms. We calculate mean value Spearman rank correlation and mutual information of each feature value and labels for all patients, the result is shown in the Figure 4.

When it comes to different patient, the efficient features will also be different. Generally speaking for all patient, the result of two methods coincide with each other. The three sensors placed in different locations and three axes can been seen as nine channels. From picture a and b, it is clear that channel1 (Horizontal forward acceleration on ankle), channel 2 (Vertical acceleration on ankle) and channel 5 (Vertical acceleration above knee) have the highest dependency with FoG symptoms. On the other hand, picture c and d shows the dominant frequency, locomotion energy and freezing index has the closest relationship with FoG among all features. According to the observation, since cross correlation and wavelet mean have quite low score and at the same time they are quite expensive in terms of computation, we decide to throw them from the list of features.

We use the threshold method to filter away features that have low score in terms of Spearman's rank correlation and mutual information separately and do classification. The classification result using new evaluation standard as quality metrics is shown in Table 2.

It is easy to find that our method performs better than the method proposed by Bachelin in all cases except patient with ID 5.

Then we can look closer to the result of detection in figure 5. The left figure shows the original method, and the right figure shows the result with new detector, it is clear than the number of false alarm decreases while most FoG windows are correctly detected. This makes sense, because under new evaluation standard, only false detection far away from real FoG period leads to a cost.

Implementation

The implementation is mainly divided into three part: Graphic user interface implemented with Qt, data training written in python and C code that running on the hardware. Since we don't have sensors that does the measurement, UART is applied to send data in window form (128*9 integer) to the hardware. The hardware we used here is the nRF52 Development kit which has 512 kB flash and 64 kB RAM memory space.

In the graphic user interface, we can choose the auto mode which uses the method mentioned before to select features, or we can manually configure the sensors numbers and lda features. The manual configuration mode is available since expert can select features that meaningful for each person individually, it may outperform the auto method which is based on data. Besides, for both mode we can also select classifiers and thresholds and relevant parameters.

After the configuration part is done, relevant data will be feed into python code as parameters and training process starts. After the training process is finished, it returns and informs the user and save results into a file at the same time.

Then user can further use GUI to send results of training and raw data measurement in window form to hardware with UART, the hardware returns the detection result back to user interface before it receives new window.

Code and more detail in found in git repository.

Conclusion and future work

The focus of this research is to detect and predict FoG events in Parkinson Disease patients through feature extraction, feature selection and decision tree. We invented a feature based on observation of raw data in time domain to exploit the periodicity characteristic of gait, it is computational less expensive in compare with other methods. A new quality metrics is proposed, it focuses more on detecting the beginning part of FoG episode and has less tolerance to false alarms in compare with traditional metrics. In general, the result using new metrics shows great improvement in compare with original method proposed by the paper about DAPHNET dataset.

As a direction for future investigation, other sensors can be applied to provide physiological information. It is also suggested to make modifications to feature selection algorithm, classification methods and classification quality metrics.

References

[1] I. Lim, E. van Wegen, C. Goede, M. Deutekom, A. Nieuwboer, A.-M. Willems, D. Jones, L. Rochester, and G. Kwakkel, "Effects of external rhythmical cueing on gait in patients with parkinson's disease: A systematic review," *Clinical rehabilitation*, vol. 19, pp. 695–713, 11 2005.

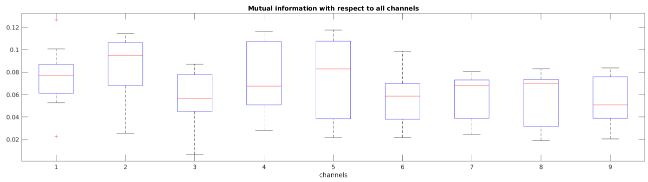
[2] A. E. Lang and A. M. Lozano, "Parkinson's disease," *New England Journal of Medicine*, vol. 339, no. 16, pp. 1130–1143, 1998. PMID: 9770561.

[3] S. Fahn, "Description of parkinson's disease as a clinical syndrome," *Annals of the New York Academy of Sciences*, vol. 991, pp. 1–14, 07 2003.

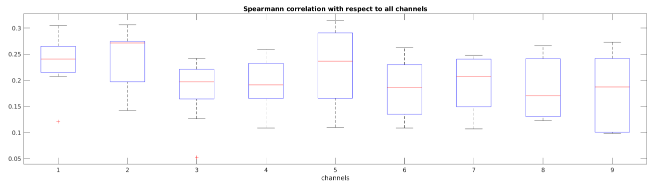
[4] R. Iansek and M. Danoudis, "Freezing of gait in parkinson's disease: Its pathophysiology and pragmatic approaches to management," *Movement Disorders Clinical Practice*, vol. 4, no. 3, p. 290—297, 2017.

[5] J. Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger, C. Peretz, and N. Giladi, "Rhythmic auditory stimulation modulates gait variability in parkinson's disease," *The European journal of neuroscience*, vol. 26, pp. 2369–75, 10 2007.

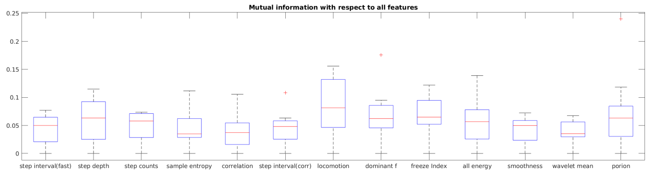
[6] S. T. Moore, H. G. MacDougall, and W. G. Ondo, "Ambulatory monitoring of freezing of gait in parkinson's disease," *Journal of neuroscience methods*, vol. 167, p. 340—348, January 2008.



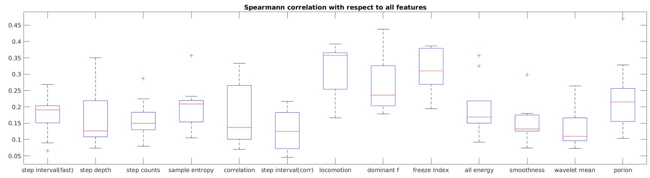
(a)



(b)



(c)

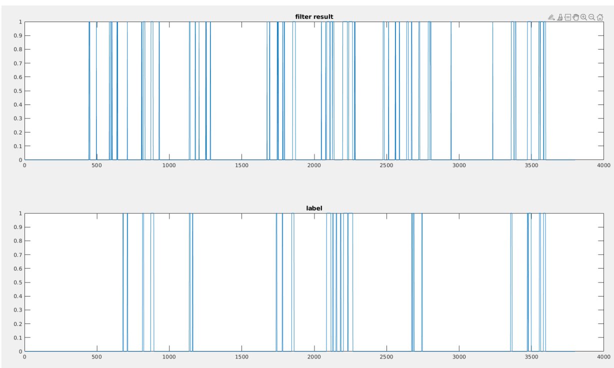


(d)

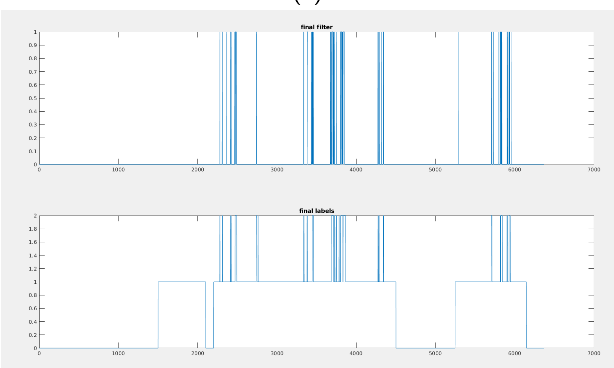
Figure 4: Figure(a) shows the mutual information between channels and FoG;Figure(b) shows the Spearmann correlation between channels and FoG; Figure(c) shows the mutual information between features and FoG; Figure(d) shows the Spearmann correlation between features and FoG;

	Result with Spearman's correlation			Result with Mutual information			Result with Bachelin's method		
ID	precision	recall	F1-score	precision	recall	F1-score	precision	recall	F1-score
1	0.8606	0.8419	0.8511	0.7808	0.8972	0.8350	0.6685	0.7470	0.7056
2	0.8854	0.9315	0.9079	0.9286	0.8191	0.8704	0.7228	0.8142	0.7658
3	0.8478	0.9733	0.9062	0.8357	0.9793	0.9018	0.8050	0.9111	0.8548
5	0.7524	0.7750	0.7635	0.7372	0.8091	0.7682	0.8161	0.8269	0.8214
6	0.7338	0.9152	0.8145	0.8232	0.8799	0.8506	0.5471	0.8763	0.6736
7	0.7704	0.7773	0.7738	0.7260	0.8389	0.7783	0.7184	0.6066	0.6578
8	0.8591	0.8861	0.8772	0.8286	0.8545	0.8413	0.7738	0.8176	0.7951
9	0.7302	0.8916	0.8028	0.9013	0.6054	0.7243	0.7294	0.7504	0.7398

Table 2: detection result



(a)



(b)

Figure 5: Figure(a) shows the detection result use Bachelin's method and its comparison with label;Figure(b) shows the detection result use Spearman's rank correlation to select features and its comparison with label;

[7] M. Bachlin, M. Plotnik, D. Roggen, I. Maidan, J. M. Hausdorff, N. Giladi, and G. Troster, "Wearable assistant for parkinson's disease patients with the freezing of gait symptom," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 436–446, 2010.

[8] F. Demrozi, R. Bacchin, S. Tamburin, M. Cristani, and G. Pravadelli, "Towards a wearable system for predicting the freezing of gait in people affected by parkinson's disease," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, pp. 1–1, 11 2019.

[9] N. K. Orphanidou, A. Hussain, R. Keight, P. Lishoa, J. Hind, and H. Al-Askar, "Predicting freezing of gait in parkinsons disease patients using machine learning," in *2018 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1–8, 2018.

[10] P. Tahafchi and J. W. Judy, "Freezing-of-gait detection using wearable-sensor technology and neural-network classifier," in *2019 IEEE SENSORS*, pp. 1–4, 2019.

[11] J. Richman and J. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American journal of physiology. Heart and circulatory physiology*, vol. 278, pp. H2039–49, 07 2000.

[12] S. Rezvanian and T. Lockhart, "Towards real-time detection of freezing of gait using wavelet transform on wireless accelerometer data," *Sensors*, vol. 16, p. 475, 04 2016.