A Robust Algorithm for Gait Cycle Segmentation

Shuo Jiang, Xingchen Wang, Maria Kyrarini, Axel Gräser
Institute of Automation
University of Bremen
Bremen, Germany
jiangs@uni-bremen.de

***Abstract*—In this paper, a robust algorithm for gait cycle segmentation is proposed based on a peak detection approach. The proposed algorithm is less influenced by noise and outliers and is capable of segmenting gait cycles from different types of gait signals recorded using different sensor systems. The presented algorithm has enhanced ability to segment gait cycles by eliminating the false peaks and interpolating the missing peaks. The variance of segmented cycles' lengths is computed as a criterion for evaluating the performance of segmentation. The proposed algorithm is tested on gait signals of patients diagnosed with Parkinson's disease collected from three databases. The segmentation results on three types of gait signals demonstrate the capability of the proposed algorithm to segment gait cycles accurately, and have achieved better performance than the original peak detection methods.**

***Keywords*—Biomedical signal processing; gait analysis; gait cycle segmentation; peak detection.**

I. INTRODUCTION

Gait analysis is the systematic study of human walking. As a newly developed research area, gait analysis has shown great potential for diagnosis in neurodegenerative diseases such as Parkinson's disease or peripheral neuropathy. Different systems are used in order to obtain gait signals, such as vision-based systems [1], goniometers [2], Inertial Measurement Units (IMUs) [3], foot pressure sensors [4]. Gait data are commonly recorded at different parts of human body, such as lower limbs, feet, and lower waist. Important gait features are usually extracted from individual gait cycle for the purpose of clinical gait assessment and monitoring [3].In order to enable gait analysis based on gait cycles, segmentation of gait cycle from a dataset containing multiple cycles is needed. One gait cycle is defined as the interval between two successive heel stride incidences [1]. The reliability of segmentation is essential for further analysis and directly influences the quality of analysis results of gait. Additionally, pathological gait tends to have more noisy data, which makes the gait cycle segmentation more challenging [5].

Several gait cycle segmentation methods have been proposed based on specific types of signal recorded from lower limbs or feet. The most commonly used segmentation algorithm is peak detection algorithm based on Z-score rule [6]. The peak detection algorithms rely on the semi-periodic characteristics of gait cycle, and the fact that the starting and ending points of a gait cycles tend ideally to be maxima or minima of the signal. Existing peak detection algorithms have achieved outstanding performance, but showed its deficiency when dealing with noisy data and outliers of gait signals.

One algorithm based on a peak detection approach was proposed by [7]. It was applied on gait signals collected from the accelerometer attached on patients' waists. Firstly, a template was found from the dataset, which corresponds to a gait cycle estimated by observing the dataset. Then the cross correlation was used to evaluate the similarity between the template and the entire acceleration signal. The peaks resulted from cross correlation could then be detected by peak detection algorithm for segmentation. The drawback of this method is that a good template must be carefully selected, and the robustness of the segmentation cannot be ensured if the dataset has large fluctuation in signal amplitude or in cycle length.

In the work presented in [4], a gait segmentation algorithm on gait signals collected from pressure sensors mounted under feet (foot-switch signals) was proposed. The starting point of each cycle was defined as the points whose difference with the subsequent point is not equal to 0, and a threshold rule was employed for gait phase segmentation. This algorithm achieved promising performance but was only evaluated on foot-switch signals. Also the choice of values for specific parameters was not discussed.

A segmentation method based on continuous wavelet transform (CWT) was proposed by [8]. The segmentation algorithm employed the peak detection algorithm on the CWT of the acceleration signals, and achieved satisfactory performance. However, the selection of proper wavelet basis is critical for the segmentation.

In the previously proposed approaches, the gait segmentation algorithms are designed for specific device configurations (sensors of different types and mounted on particular positions of body). They highly depend on the type of employed sensor measurement system and the characteristic of signal. Therefore, in the presented work, a robust algorithm for gait cycle segmentation is developed, its segmentation result is less influenced by noise and outliers, and it is capable of segmenting gait cycles in different types of gait signals. This algorithm does not focus on specific physical meaning represented by the signal itself, but on the statistical features depicting how the gait cycles are uniformly distributed.

This paper is organized as follows: In section II, the improved algorithm is introduced step by step. Section III presents the performance comparison of the proposed algorithm and the original peak detection algorithm [9] on three different types of common gait signals and section IV discusses the conclusion and future work.

II. GAIT SEGMENTATION ALGORITHM

ISBN 978-0-9928626-7-1 © EURASIP 2017

The proposed segmentation algorithm is based on the peak detection algorithm provided in [9].

The original algorithm presented in [9] scans through the whole segmentation reference dataset from the beginning to the end. A threshold T is introduced as a standard of extreme detection. Firstly, the first maximum of the dataset is determined, and then the algorithm searches for the next minimum. This minimum is the closest point to the first maximum where the difference in amplitude between the maximum and this point is larger than the threshold T. Next the algorithm searches the next maximum, which is the closest point to the last minimum detected and that the difference in amplitude between this point and the last maximum is larger than the threshold T. Furthermore, the algorithm switches between maxima search and minima search to determine the location of maxima and minima according to the principle that a maximum is always between two minima, and vice versa.

However the original algorithm cannot deal with situations where a peak has very large amplitude. Also if the peak has inadequate amplitude, the original peak detection algorithm is not able to detect it. The situation where a peak with amplitude over 50% smaller than the others is not detected is shown in Fig. 1

Inour enhanced algorithm, an optimal threshold value as in original algorithm should be determined, which describes the difference between the current peak and its surrounding values. The rule of choosing an optimal threshold is that the optimal threshold should minimize the variance of lengths of all detected cycles. When several sensors record synchronously, one of the signals that provide the minimum variance of cycle lengths is chosen to be the reference signal for segmentation. The peak detection algorithm introduced in [9] is applied on the reference signal with the determined optimal threshold. The detected peaks were all considered as potential candidates of starting and ending points of gait cycles. Then the false peaks are eliminated and missing peaks are interpolatedat both sides of the determined initial cycle to obtain the final results.

The algorithm can be divided into five steps:

a) Find the optimal threshold value for peak detection:

In peak detection algorithm [9], the threshold as introduced should be carefully assigned with an optimal value. With the normalization on provided dataset, the value of T is a positive non-zero value smaller than1. Usually $T = 0.5$ is chosen for normalized data and achieves satisfactory performance [4] (for this experiment, the datasets are normalized from 0 to 1 in magnitude). But it may not always be the optimal segmentation value

Figure 1. Segmentation result of original peak detection algorithm. One peak is not detected due to inadequate amplitude.

based on the characteristic of the dataset. For a specific threshold T, the result of original peak detection algorithm is a vector $V(i)$ where i=1,2,3,...,n denoting the position of peaks in sample domain. n is the total number of detected peaks. The distance d_i between peaks is defined as

$$d_i = V(i+1) - V(i) \qquad i = 1,\dots,n-1 \tag{1}$$

One cycle is defined as the fragment between two peaks and the average cycle length $\bar{d}$ computed using all detected peaks is defined as

$$\bar{d} = \frac{1}{n-1}\sum_{i=1}^{n-1} d_i \tag{2}$$

The variance of cycle length (var) is defined as

$$var = \frac{1}{n-1}\sum_{i=1}^{n-1} (d_i - \bar{d})^2 \tag{3}$$

The criterion for the determination of the optimal threshold $\hat{T}$ selection is

$$\hat{T} = \underset{T}{\operatorname{argmin}}\left(var\left(d(T)\right)\right) \tag{4}$$

$\hat{T}$ represents the estimation of the optimal threshold. This optimal threshold can be determined with the following procedures: increase T from T_{low} to T_{high} with a step t in each iteration. Calculate the variance of cycle length using (3) in every iteration and choose $\hat{T}$ to be the T with minimum variance of cycle length. The values in the presented approach for T_{low} is 0.2, for T_{high} is 0.8 and $t$is 0.01. The reason why T increases not from 0 to 1 but from 0.2 to 0.8 is that, if T is too big (bigger than 0.8), only very few points can be detected as peaks. Consider an extreme situation where only two peaks are detected, the variance of cycle length is zero. In this case the optimal threshold cannot be determined due to small number of detected peaks. When threshold is too small (smaller than 0.2), another extreme situation will happen, namely that all points are detected as peaks, and the variance of cycle length is also zero, which will lead to a false selection of threshold as well. The relationship between the threshold and the variances of cycle length from one example gait signal is depicted in Fig.2.

b) Find a dataset to be the segmentation reference:

This step is to select the most reasonable dataset to be the segmentation reference dataset, when several synchronous datasets are provided. The principle of this step is to select the reference dataset which is less affected by noise and potentially gives the most promising segmentation result as the reference for performing the segmentation. The selecting criterion is given by (5). Segmentation on the optimal reference dataset should minimize the variance of cycle length.

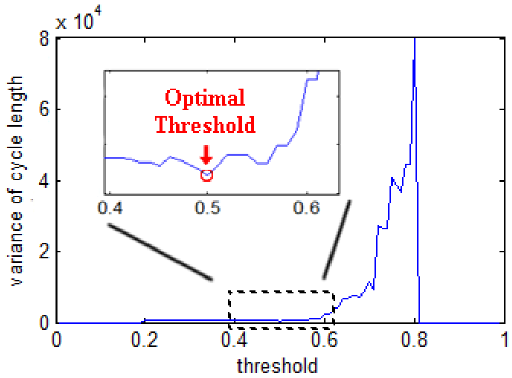

Figure 2. Determination of the optimal threshold by minimizing the variance of cycle length.

$$data\_opt = \underset{data}{\operatorname{argmin}}\big(var(data\ j)\big) \tag{5}$$

where $data\ j$ represents one of multiple provided synchronous dataset. Afterwards, the peak detection algorithm is applied on the reference signal with the optimal threshold $\hat{T}$, which is determined in step a). The heel strike incidences were considered as the starting and ending points of a gait cycle as defined in previous research [1]. Minima are detected to be the segmentation flags, where data fragment between two flags is one complete gait cycle [10]. However when the dataset has large fluctuation or significant outliers, false detection can occur.

c) Find the initial potential cycle (IPC) in reference dataset:

Two consecutive peaks should be selected in this step as the initial cycle of the algorithm. This IPC should fulfill the criterion(6)

$$0.9\bar{d} < length(IPC) < 1.1\bar{d} \tag{6}$$

With this criterion,the tolerance of the fluctuation of the cycle is +-10%. When it is fulfilled, it is highly credible that this cycle (IPC) is a correct segmentation, as shown in Fig .3. If more than one cycle fulfills the criterion (6), IPC is arbitrarily chosen.

d) Count the potential cycle segmentations on both sides of the IPC:

As the initial credible cycle has been found, the proposed algorithm will deal with the peaks on both sides of the IPC in this step. Each peak is assigned with an initial state $S(\mathrm{i})$ which will be updated afterwards. The definition of the states is presented in (7).

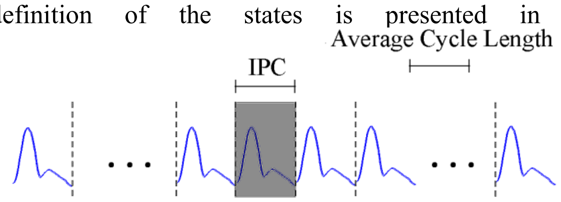

Figure 3. Find the initial potential cycle (IPC)

$$S(\mathrm{i})\begin{cases} =0 & invalid \\ =-1 & unknown \\ \geq 1 & one\ or\ more\ hidden\ period \end{cases} \tag{7}$$

At the beginning, only the IPC is set as 1(valid), and all other peaks are set as -1 (unknown). Then a loop of index i (for the i^{th} peak in the peak list) checks the peaks one by one from the IPC to each side, using the following criteria to update the states of peaks.

I) if $0.9\bar{d} < d_i < 1.1\bar{d}$, set $S(i) = 1$

II) if $d_i < 0.9\bar{d}$, set $S(\mathrm{i}) = 0$

III) else calculate $p(i) = V(i) - V(m)$, where m is the position of last non-zero peak, if $p(\mathrm{i}) > 1.1\bar{d}$, then do the following:

$$\begin{cases} \text{if} \begin{cases} 0.9 \times round(\frac{p(\mathrm{i})}{\bar{d}}) \times \bar{d} < p(\mathrm{i}) \\ and \\ p(\mathrm{i}) < 1.1 \times round(\frac{p(\mathrm{i})}{\bar{d}}) \times \bar{d} \end{cases} \\ \text{then set } S(\mathrm{i}) = round(\frac{p(\mathrm{i})}{\bar{d}}), \\ \text{else set } S(\mathrm{i}) = 0 \end{cases}$$

$p(\mathrm{i})$ is a temporary variable defined as the distance between two extreme values which potentially has one or more cycles in between. $V(i)$ is defined as the same in (1).

The criterion I indicates the case that a cycle is detected. The criterion II indicate the situation that when the distance between two peaks is too small (smaller than 0.9 time of the average cycle), we assume there is no complete cycle in between. It may be caused by noise. The idea of the criterion III is that when the distance between two peaks is larger than 1.1 times of the average cycle, hidden cycles (missing cycles) exist between the two peaks. The number of hidden cycles is calculated with (8)

$$number = round(\frac{p(\mathrm{i})}{\bar{d}}) \tag{8}$$

If $p(i)$, $i=1,\ldots,n$, exceeds the 10% tolerable region of $number \times \bar{d}$, the unknown point is invalid. For searching the cycles after IPC, the i increases and for searching the cycles before IPC, the i decreases. That is the reason the last non-zero is used in criterion III. There are cases where there are one or more peaks in one cycle, which are false peaks introduced by noise.

e) Eliminate and interpolate peaks for both sides.

In step *(d)*, each peak is set with a state of 0 (invalid), or$\geq$1(one or more cycles on right side). In this step, the proposed algorithm eliminates the invalid peaks and interpolates the hidden peaks. A loop from IPC to the end of signal is set to eliminate and interpolate peaks on the right side of IPC, and another loop from IPC to the

ISBN 978-0-9928626-7-1 © EURASIP 2017

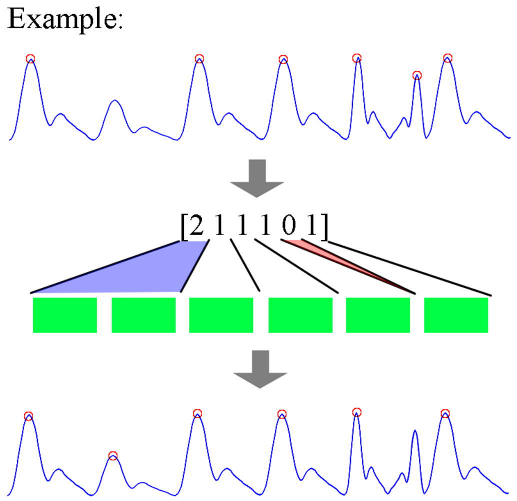

Figure 4.Example of segmentation result using improved algorithm

beginning of the signal is set to process the left side of signal. In this paper, linear interpolation is used, which means that if more than one cycle exist between two peaks, new peaks will be interpolated with the same distance in between.

In Fig. 4, an example of the gait segmentation results using the proposed algorithm is shown. The upper figure shows the signal whose peaksare detected by using only the original peak detection algorithm, and marked with red circles. the lower figure shows the peak detection with the proposed method. The numbers in the vector represents the number of existing potential cycles between two peaks. For example, between the first and second peak marked with the red circle, the proposed algorithm detects 2 cycles. Then the proposed method interpolates one peak in between. Moreover, in same figure the fifth peak is invalid, since the proposed method gave 0 as state and this peak is a false detection and therefore is eliminated.

III. PERFORMANCE ANALYSIS

In the performance analysis of the proposed algorithm, three databases of Parkinson's disease gait signals were used to test the segmentation performance, consisting of three typical gait measurements.

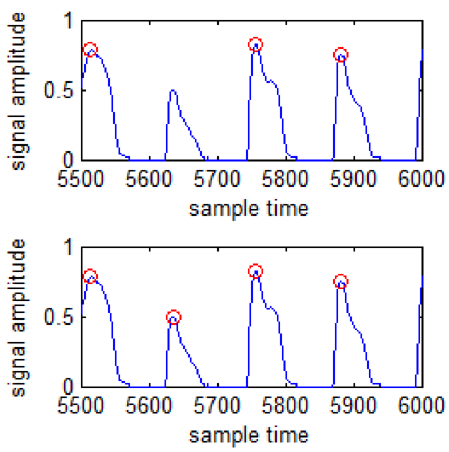

Figure5.(a): Segmentation on vertical ground reaction force signal with original peak detection algorithm. (b): Segmentation with proposed algorithm. Red circles are the detected peaks, which represent the starting and ending points of segmented gait cycles

The first database is open source gait signal database on the website of Physionet [11].The data used for test are collected at sampling rate of 100 Hz. It records the vertical ground reaction force by 8 pressure sensors under both feet. The original peak detection algorithm and proposed algorithm are applied and the performances are compared. The result is shown in Fig.5. As can be seen, the peak is not successfully detected in the second cycle using only original peak detection algorithm, but the improved algorithm is capable to detect this point by interpolating the missing peak.

The second database comes from the data of gait signal [3] from Institute of Automation, University of Bremen. In this test, thesignal vector magnitude of acceleration gait signal with sampling rate of 30 Hz is segmented [10]. The signal vector magnitude is defined in (9).

$$svm_{acc} = \sqrt{a_x^2(k) + a_y^2(k) + a_z^2(k)} \tag{9}$$

a_x, a_y and a_z are acceleration signals in three axes collected by an inertial measurement unit (IMU) on the waist of patient and k is sample number. The segmentation results are shown in Fig. 6.

The Fig.6(a) shows the results of the original peak detection algorithm which detected 12 peaks. Peak around sample time 1480 is not detected due to its inadequate amplitude. Around sample time 2130, two peaks are detected with very short distance to each other, however the one on the right is false detection introduced by noise. In Fig. 6(b), the proposed algorithm is able to detect thepeak around sample time 1480, and the false peak around time sample 2130 is eliminated successfully.

The third database comes from the data analysis of gait signal from Institute of Automation, University of Bremen, which contains the measurement from Inertial Measurement Unit (IMU)

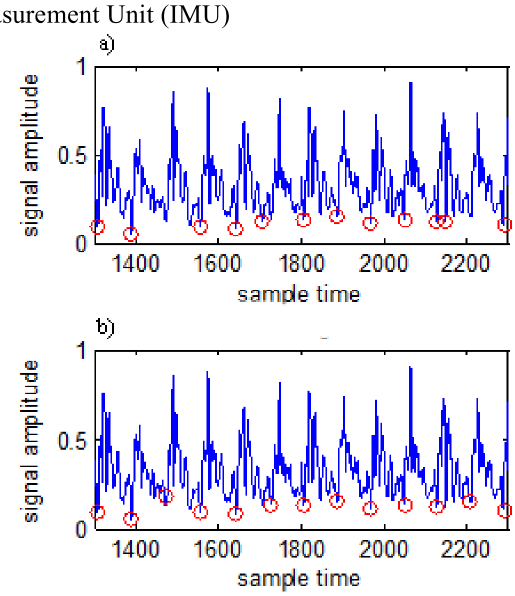

Figure 6.(a): Segmentation on signal vector magnitudewith original peak detection algorithm. (b): Segmentation withproposed algorithm.

ISBN 978-0-9928626-7-1 © EURASIP 2017

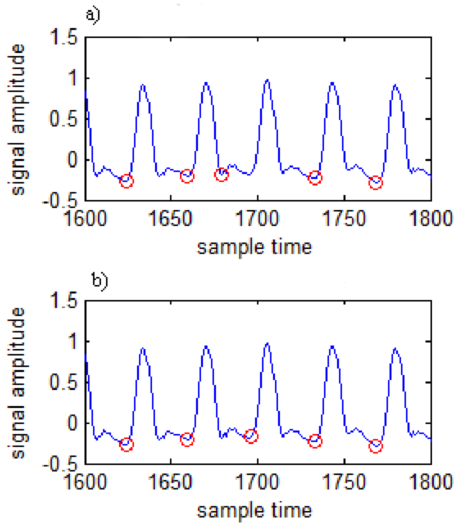

Figure 7. (a): Segmentation on knee joint angle signal with original peak detection algorithm. (b): Segmentation with proposed algorithm.

of the hip and knee joint angles of patients with Parkinson's disease [12]. The analyzed data is the knee joint angle signal at sampling rate of 30 Hz. In Fig. 7(a), a false peak is detected around time sample 1680 by original peak detection algorithm. This false peak can be eliminated by the proposed algorithm, and the peak with correct location is detected, as can be seen in Fig. 7 (b).

Finally, a comparison of the variance of cycle length for all the three datasets is performed for the evaluation of the proposed method, and the results are presented in Table I. The reason the variance of cycle length is used to evaluate the performance is that when false or missing peaks exist, the variance of cycle length would be large. The proposed algorithm yields smaller variance of cycle length, which means that the cycle lengths as segmented by proposed method are more uniformly distributed compared to the result of original peak detection algorithm. This is because the proposed algorithm is able to eliminate false peaks and detect missing peaks, which avoids large variance of cycle length.

TABLE I. COMPARISON OF ALGORITHMS ON THREE DATA CATEGORIES

Measurement method	**Variance of cycle length using original Peak detection algorithm**	**Variance of cycle length using proposed algorithm**
Foot-switch	215.6	18.9
Signal vector magnitude by IMU	831.7	10.7
Knee angle by IMU	21.4	1.6

IV. CONCLUSION

In this paper, a robust gait segmentation algorithm is proposed based on a peak detection approach. The criterion of segmentation is to minimize the variance of cycle length by finding an optimal threshold for peak detection. The algorithm employs elimination and interpolation in time domain to remove false peaks and detect missing peaks to improve the segmentation effectiveness due to the fact that the cycle lengths tend to have uniform distribution. The performance of the proposed improved algorithm is evaluated on three different types of datasets, i.e., the foot pressure data, the acceleration data and the knee joint angle trajectories. The results of all three types of data have shown promising results with all incorrect peaks eliminated and missing peaks interpolated, while the results obtained using only original method contain detection errors. Additionally, the performance of the proposed algorithm is further evaluated using the variance of cycle lengths.

The limitation of the algorithm is that the algorithm was not yet fully evaluated with data containing large fluctuation and non-periodic behavior. Enlarging the tolerance given in (6) is a compromising solution to small range non-periodic behavior, and this adaptation should be further investigated in the future work.

REFERENCE

[1] Saravana K. Natarajan, Xingchen Wang, Matthias Spranger, and Axel Gräser, "Reha@Home-A Vision Based Markerless Gait Analysis System for Rehabilitation at Home," in proceeding of IEEE the 13th IASTED international conference on Biomedical Engineering. 2017

[2] Angela Shiratsu, H.J.C.G. Coury, "Reliability and accuracy of differentsensors of a flexible electrogoniometer," Clinical Biomechanics, vol. 18,pp. 682-684, 2003

[3] Xingchen Wang, Maria Kyrarini, DanijelaRistić-Durrant, Matthias Spranger, and Axel Gräser, "Monitoring of gait performance using Dynamic Time Warping on IMU-sensor data,"IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 1-6,2016

[4] Valentina Agostini, Gabriella Balestra, Member, IEEE, and Marco Knaflitz, Member, IEEE, "Segmentation and classification of gait cycles," IEEE Transactions on Neural Systems and Rehabilitation Engineering, VOL. 22, NO. 5, September 2014

[5] Cornelius Glackin, Christoph Salge, Daniel Polani, Markus Tüttemann, Carsten Vogel, Carlos Rodriguez Guerrero, Victor Grosu, Svetlana Grosu, Andrej Olenšek, MatjažZadravec, ImreCikajlo, Zlatko Matjačić, Adrian Leu, and DanijelaRistić-Durrant," Learning gait by therapist demonstration for natural-like walking with the CORBYS powered orthosis," 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)

[6] E. Kreyszig (1979). "Advanced Engineering Mathematics (Fourth ed.),"Wiley.p. 880, eq. 5. ISBN 0-471-02140-7.

[7] Mitsuru Yoneyama, Yosuke Kurihara, Kajiro Watanabe, and Hiroshi Mitoma, "Accelerometry-based gait analysis and its application to Parkinson's Disease assessment," IEEE Translations on Neural Systems and Rehabilitation Engineering, VOL. 22, NO.3, May2014.

[8] Mohamed Boutaayamou, Olivier Brüls, Vincent Denoël, Cédric Schwartz, Marie Demonceau, GaëtanGarraux, and Jacques G. Verly, "Segmentation of gait cycles using foot-mounted 3D accelerometers,"2015 International Conference on 3D Imaging (IC3D)

[9] PeakDet: [Avaliable Online]
http://www.billauer.co.il/peakdet.html

[10]Yu-Liang Hsu, Pau-Choo (Julia) Chung, Fellow, IEEE, Wei-Hsin Wang, Ming-ChyiPai, Chun-Yao Wang, Chien-Wen Lin, Hao-Li Wu, and Jeen-Shing Wang, Member, IEEE, "Gait and balance analysis for patients with Alzheimer's Disease using an inertial-sensor-based wearable instrument," IEEE Journal of Biomedical and Health Informatics, VOL. 18, NO. 6, November 2014

[11] PhysioBank Databases, Gait and Balance Databases, Gait in Parkinson's Disease, dual tasking in PD; Eur J Neuro, 2005.

[12] Maria Kyrarini, Xingchen Wang, and Axel Gräser, "Comparison of vision-based and sensor-based systems for joint angle gait analysis," 2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA). pp. 375-379.