Recognizing Parkinsonian Gait Pattern by Exploiting Fine-Grained Movement Function Features

TIANBEN WANG and ZHU WANG, School of Computer Science,
Northwestern Polytechnical University
DAQING ZHANG, Institut Mines-Télécom/Télécom SudParis
TAO GU, RMIT University
HONGBO NI, JIANGBO JIA, and XINGSHE ZHOU, School of Computer Science,
Northwestern Polytechnical University
JING LV, Zhuhai Kingsoft Office Software Co., Ltd

Parkinson's disease (PD) is one of the typical movement disorder diseases among elderly people, which has a serious impact on their daily lives. In this article, we propose a novel computation framework to recognize gait patterns in patients with PD. The key idea of our approach is to distinguish gait patterns in PD patients from healthy individuals by accurately extracting gait features that capture all three aspects of movement functions, that is, stability, symmetry, and harmony. The proposed framework contains three steps: gait phase discrimination, feature extraction and selection, and pattern classification. In the first step, we put forward a sliding window–based method to discriminate four gait phases from plantar pressure data. Based on the gait phases, we extract and select gait features that characterize stability, symmetry, and harmony of movement functions. Finally, we recognize PD gait patterns by applying a hybrid classification model. We evaluate the framework using an open dataset that contains real plantar pressure data of 93 PD patients and 72 healthy individuals. Experimental results demonstrate that our framework significantly outperforms the four baseline approaches.

CCS Concepts: • **Applied computing** → **Health care information systems**; *Health informatics*

Additional Key Words and Phrases: Parkinson's disease, gait phases, gait stability, gait symmetry, gait harmony, gait pattern recognition

ACM Reference Format:
Tianben Wang, Zhu Wang, Daqing Zhang, Tao Gu, Hongbo Ni, Jiangbo Jia, Xingshe Zhou, and Jing LV. 2016. Recognizing parkinsonian gait pattern by exploiting fine-grained movement function features. ACM Trans. Intell. Syst. Technol. 8, 1, Article 6 (August 2016), 22 pages.
DOI: http://dx.doi.org/10.1145/2890511

6

This work is supported in part by the National Natural Science Foundation of China (No. 61332013, 61402369), the National Key Research and Development Program of China (No. 2016YFB1001400), the Natural Science Foundation of Shaanxi Province (No. 2015JQ6237), and the Fundamental Research Funds for Central Universities (No. 3102014JSJ0004).
Authors' addresses: T. Wang, Z. Wang, H. Ni, J. Jia, and X. Zhou, School of Computer Science, Northwestern Polytechnical University, Xi'an Shaanxi, 710072, China; email: wangtianbengx@163.com, wangzhu@nwpu.edu.cn, nihb@nwpu.edu.cn, jiajiangbo758@gmail.com, zhouxs@nwpu.edu.cn; D. Zhang, SAMOVAR, Télécom Sud-Paris, CNRS, Université Paris-Saclay, 9 rue Charles Fourier 91011 Evry, France; email: daqing.zhang@telecom-sudparis.eu; T. Gu, School of Computer Science and IT, RMIT University, Melbourne VIC 3001 Australia; email: tao.gu@rmit.edu.au; J. Lv, Zhuhai Kingsoft Office Software Co., Ltd, Zhuhai Guangdong, 519015, China; email: wings6067@126.com.

DOI: http://dx.doi.org/10.1145/2890511

1. INTRODUCTION

The world population is rapidly aging. With this change occurring worldwide, health issues associated with an aging population are increasing. The movement disorder Parkinson's disease (PD) is one of the common diseases among elderly people, which has a serious impact on their daily lives [Zhang et al. 2005]. PD is caused by a degenerative disorder of the central nervous system of a human body, leading to the gradual loss of body movement function, along with anxiety and depression. According to a report from World Health Organization [Gupta and Bala 2013; Michael J. Fox Foundation 2016], there are 6.3 million people suffering from PD worldwide, and the number in China alone is more than 1.7 million, which is about 1% of the population over age 60. With an increasing amount of research focusing on diagnosing PD, two research issues are of paramount importance: (1) How can we prediagnose PD and help doctors make early interventions to delay its occurrence or even prevent it, especially when PD symptoms are not apparent? (2) How can we assess the levels of symptoms and assist doctors to make treatments to avoid falling into worse conditions?

Among many recent efforts, quantitative gait analysis has been a promising research direction, as it intends to reveal gait variation [Iosa et al. 2012, 2013, 2014; Yogev et al. 2007; Sant'Anna et al. 2011] and further study the impact of specific diseases on gait performance [Iosa et al. 2012, 2014; Yogev et al. 2007; Baltadjieva et al. 2006; Hausdorff et al. 2007; Yogev et al. 2005; Frenkel-Toledo et al. 2005]. It provides a new assistive tool for health professionals to address the two aforementioned issues through recognizing PD gait patterns. To recognize PD gait patterns, a popular approach [Yogev et al. 2007; Wu and Krishnan 2010; Hausdorff et al. 2000] is first to extract the stride-to-stride time series (consisting of the duration of each gait cycle, in which a gait cycle refers to the process from one heel strike to the next heel strike of the same foot), then extract gait features such as the similarity between two feet [Yogev et al. 2007], standard deviation and Signal Turns Count (STC) [Wu and Krishnan 2010], average values of stride time and speed, fluctuation magnitude, and fluctuation dynamics [18]. Alternatively, a symbolization method [Sant'Anna et al. 2011] has been proposed to transform the acceleration signal of each foot into a symbolic sequence; the similarity between two feet is calculated by comparing the symbolic sequence of two feet. Finally, a classifier is used to recognize PD gait patterns.

The traditional approaches generally work well when the symptoms of PD patients are apparent. However, when symptoms are not apparent, they exhibit two shortcomings. First, the stride-to-stride time series in the traditional methods is typically coarse-grained, and they ignore all the detailed information hidden in each gait phase. In fact, a complete gait cycle (i.e., two adjacent strides) contains four gait phases—swing phase, initial double-support phase, single-support phase, and end double-support phase [Iosa et al. 2013]. The information hidden in each gait phase is important for gait analysis, such information including the variation of each gait phase, the similarities of each gait phase between two feet, and the rates between different gait phases. However, it is impossible to obtain this information from stride-to-stride time series. Second, the traditional approaches recognize PD gait patterns by extracting one aspect of gait features. This may not achieve the desired recognition accuracy since several works [Michael J. Fox Foundation 2016; National Institute of Neurological Disorders and Stroke 2016; Iosa et al. 2014] demonstrate that PD actually affects several aspects of gait simultaneously, such as gait stability, gait symmetry, and gait harmony.

To address the shortcomings in the traditional approaches, in this article, we first obtain the four complete gait phases from gait plantar pressure data, making it possible to compute fine-grained gait features. Based on these gait phases, we then extract gait features from three aspects of movement functions—gait stability, gait symmetry, and gait harmony—from which we can understand gaits more comprehensively and

recognize PD gait patterns efficiently. However, this is not a trivial task for the following reasons. First, we obtain fine-grained gait features on each gait phase level rather than on stride-to-stride level. However, sensor data are inherently noisy. As a result, plantar pressure data may contain much noise, making it difficult to extract gait phases accurately. Second, it is challenging to normalize the gait features that can characterize the three movement functions, that is, gait stability, gait symmetry, and gait harmony based on plantar pressure data only. Third, even though two subjects belong to the same gait pattern group, each of their gait features may be significantly different as different subjects have different weights and walking speeds. Subject discrimination is one of the key obstacles in gait pattern recognition.

To tackle these challenges, we propose a computation framework that leverages the fine-grained gait features to capture multiple aspects of gait phases, thus recognizes gait patterns more efficiently. The contributions of this work are summarized as follows.

—To discriminate four gait phases, we propose a sliding window–based model to identify Heel Strike and Toe Off events, then recognize four gait phases: swing phase, initial double-support phase, single-support phase, and end double-support phase. The proposed model is able to adapt its behavior to noisy plantar pressure data.
—We extract fine-grained gait features from all three aspects of movement functions—stability, symmetry, and harmony—by computing relative variation, similarity between two feet, and special rates of different gait phases, respectively. These three aspects of gait features correspond to the typical symptoms of PD: gait instability, asymmetry, and inharmony, respectively. In addition, all features are normalized to minimize individual differences and increase their comparability. Experimental results demonstrate that PD patients show much weaker stability, symmetry, and harmony than healthy individuals.
—To achieve high recognition accuracy, we construct a hybrid classification model using the Adaboost method (i.e., a kind of Ensemble Method), which contains several BP neural network classification models. We evaluate the proposed framework using an open dataset that contains real plantar pressure data of 93 PD patients and 72 healthy people. Experimental results demonstrate that our framework outperforms all the baseline methods in terms of Area Under ROC Curve (AUC), precision and recall.

The rest of the article is organized as follows. We first review the related work in Section 2. We then formally define the problem in Section 3 and present our computation framework in Section 4, followed by the evaluation results in Section 5. Finally, we present our conclusions and discuss future work in Section 6.

2. RELATED WORK

In this section, we briefly review the related work, which can be grouped into three categories. The first line of research is on gait-phase discrimination. Senanayake and Senanayake [2010] propose a gait-phase detection algorithm that applies fuzzy logic on kinematic and kinetic parameters. The algorithm is embedded in a real-time data-acquisition system, which consists of four force-sensitive resistors and two inertial sensors to obtain foot-pressure patterns and knee flexion/extension angle. Han et al. [2009] demonstrate an adaptive window-based gait-phase discrimination method based on three-axis accelerations of the ankle during walking. The basic idea of this method is to use four narrow windows to detect initial contact (IC) point and end contact (EC) point. Kong and Tomizuka [2008] propose a fuzzy logic–based gait-phase discrimination method on ground contact forces. Four pressure sensors are located under the hallux, the first metatarsal, the fourth metatarsal, and heel, respectively, to collect pressure data and then apply fuzzy logic to overcome the weakness of threshold-based method,

that is, poor availability for smooth and continuous signal changes. Our work is motivated by these research efforts in the way that we leverage gait phases to compute fine-grained gait features. Our previous work [Wang et al. 2015] discriminates the gait phases from plantar pressure data by identifying Heel Strike (HS) and Toe Off (TO) events. In contrast, in this article, we proposed a sliding-window model to detect the gait phases, which achieves better performance when handling noisy plantar pressure data.

The second category focuses on gait feature extraction. Iosa et al. [2014] extract gait features from upper-body accelerations to access gait stability, harmony, and symmetry, and finally identify which feature is directly related to severity of the impairment. They conduct further research on the stability and harmony of gait in children with cerebral palsy [Iosa et al. 2012]. These methods perform on upper trunk acceleration and do not use gait phases. Iosa et al. [2013] also investigate the relationship between the proportions of gait phases and the golden ratio ($\approx$1.618034) based on computer vision techniques. They find that the ratio between the entire gait cycle and standing phase and the ratio between standing and the swing phase are both approximately equal to the golden ratio. Sant'Anna et al. [2013] propose a novel method to access gait symmetry, which transforms inertial sensor signals into a sequence of symbols, then measures the time interval between two consecutive segments of the same symbol for two feet. Finally, the total difference between two feet can be calculated. However, it cannot directly adapt to the plantar pressure data; since the pressure value keeps to zero in the whole swing phase, we cannot segment the signal into equiprobable quantization. Moreover, another classical and popular way to extract gait features is to measure the fluctuation of stride-to-stride time [Yogev et al. 2007; Baltadjieva et al. 2006; Wu and Krishnan 2010; Hausdorff et al. 2000]; the feature based on stride-to-stride time is too coarse-grained since it ignores the information between gait phases. In contrast, we extract features based on all the gait phases, which are generated from plantar pressure data. These gait features reflect three aspects of movement functions: stability, symmetry and harmony, respectively.

The third line of research focuses on recognizing different gait patterns. Gandomkar and Bahrami 2012 propose a markerless vision-based technique to extract gait features from human walking sequences, and finally differentiate normal and abnormal gaits. Sant'Anna et al. [2011] further use the extracted gait feature, SI_{symb}, to classify gait patterns corresponding to healthy people and PD patients with initial symptoms, that is, affecting one side more than other. Similarly, Wu and Krishnan [2010] extract two gait features—standard deviation of stride interval (σ_r) and STC of stride interval, then classify gait patterns corresponding to healthy people and PD patients by employing a least squares support vector machine (LS-SVM) algorithm. However, features σ_r and STC both indicate only the stability of gait rhythm. Due to the complexity of PD symptoms and individual differences, it typically ends up with poor performance if we classify gait patterns based only on features reflecting a single aspect of movement function and linear classifier. To address this issue, we propose classifying gait patterns based on fine-grained features that characterize all three aspects of movement functions. Compared with our previous work [Wang et al. 2015], which applies the naïve BP neural network to model the extracted features, in this article, we construct a hybrid classification model by composing a set of BP neural network models.

3. PROBLEM STATEMENT

In this article, we mainly focus on the pressure data at toes and heels. Specifically, the pressure data collected from sensors placed under toes and heels can be regarded as a time series, which is formally described as follows:

$$TS = \{s_1, s_2, ..., s_n\},$$

where n is the number of recordings, and s_i $(1 \leq i \leq \text{n})$ is a sampling point that is denoted as:

$$s_i = (t_i, p_{T\_i}, p_{H\_i}),$$

where t_i is the timestamp, and p_{T_i} and p_{H_i} are the pressure value at the toe and the heel, respectively.

Given the gait-pressure data of an individual (TS_l, TS_r), that is, gait pressures of both the left and right foot, we can measure the following motion characteristics of one's gait.

—*Stability*: It reflects not only the steadiness of gait rhythm but also the steadiness of pressure value at toes and heels of two feet. The normal gait should have good stability. Meanwhile, it is well known that postural instability is one of the most important symptoms of PD [Iosa et al. 2012, 2014; Yogev et al. 2007; Wu and Krishnan 2010].

—*Symmetry*: It reveals the similarity of two feet, which contains the similarity between the gait rhythm and the pressure value of two feet. High symmetry between two feet is an important metaphor of health. The symptoms of PD usually appear only on one side of the body at the early-to-mid stage, which usually result in severe gait asymmetry.

—*Harmony*: It characterizes the percentage of different gait phases. For example, while the swing phase of a healthy gait should account for about 40% of the whole gait cycle, that of the gait corresponding to Parkinson's disease usually has a significantly lower percentage [Iosa et al. 2012, 2013]. The reason is that Parkinson's patients need to increase the standing phase to keep balance.

Per this discussion, gait stability, symmetry, and harmony characterize people's movement functions from different aspects, which can be estimated based on gait features. We aim to classify different gait patterns corresponding to PD patients and healthy individuals by combining these features. We formally define the problem as follows.

Problem. Given a time series of two feet gait pressure $TSS = \{(TS_{l\_1}, TS_{r\_1}), (TS_{l\_2}, TS_{r\_2}), \ldots\ldots, (TS_{l\_n}, TS_{r\_n})\}$, where each pair comes from either a healthy individual or a PD patient, the goal is to extract features depicting the movement functions, then correctly identify the gait patterns based on these features.

4. METHOD

In this section, we present the detailed design of the proposed gait analysis framework.

4.1. Gait Analysis Framework

Gait is a complex process, which consists of several phases. Discriminating different gait phases is an important aspect of analyzing the gait process, and is the basic procedure to quantitatively characterize gait and do further analysis. Therefore, in this article, we first discriminate four gait phases (swing, initial double standing, single standing, end double standing) from the foot pressure time series of both feet during walking. The description of these four gait phases is provided in Section 4.2. Based on these gait phases, we can further extract and select valid features to quantitatively describe gait from the aspects of stability, symmetry and harmony. Finally, in order to demonstrate the effectiveness of these features, we use them to classify gait patterns corresponding to PD patients and healthy individuals by using a hybrid classification model. Figure 1 shows the overview of the proposed gait analysis framework, which

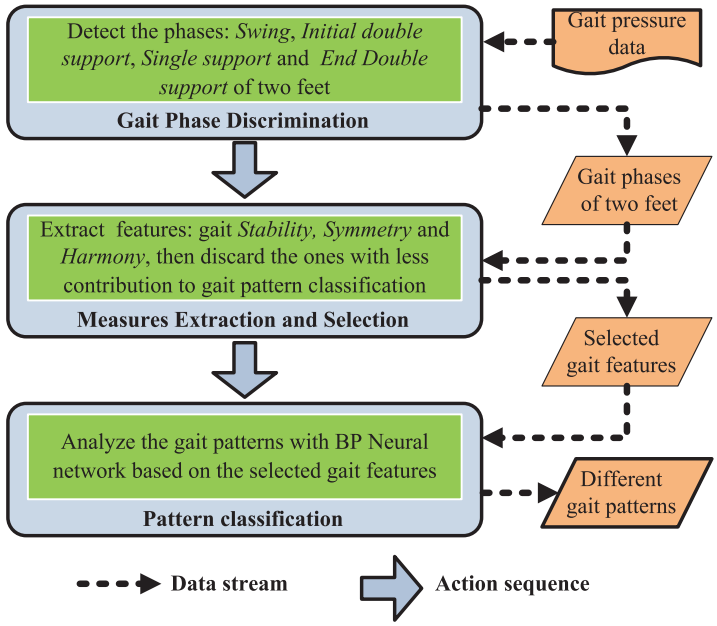

Fig. 1. Overview of gait analysis framework.

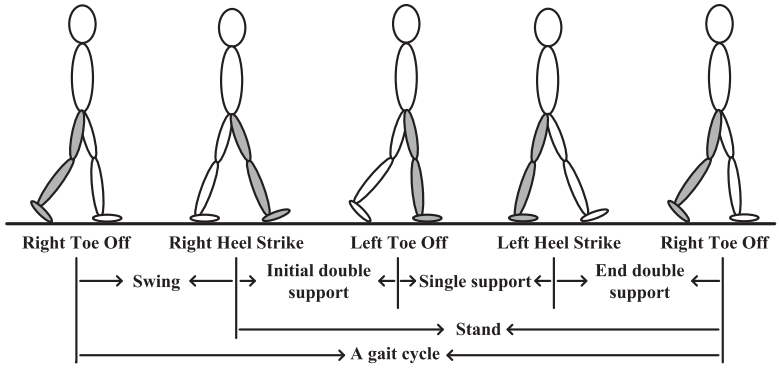

Fig. 2. Four gait phases and transitions between them. The objective foot is highlighted in gray.

includes three steps: gait phase discrimination, features extraction and selection, and pattern classification. The details of these steps are presented in the following sections.

4.2. Gait Phase Discrimination

As shown in Figure 1, given a series of gait data, the first task is to discriminate different gait phases. Gait-phase discrimination is an effective tool to analyze gait and has been widely used in gait analysis [Iosa et al. 2013; Senanayake and Senanayake 2010; Han et al. 2009; Kong et al. 2008]. In the proposed framework, we adopt the four-phase criterion [Iosa et al. 2013], which segments one gait cycle into four phases: swing phase, initial double-support phase, single-support phase, and end double-support phase. Figure 2 depicts these four phases and the transition between each of them. It is clear that the four phases form a circulation and appear alternately during walking. In contrast, the initial double-support phase and end double-support phase will disappear

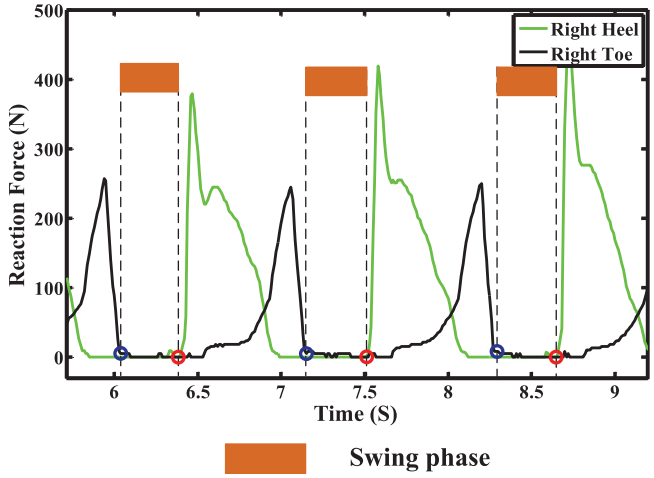

Fig. 3. Swing phases with irregular fluctuations. The blue and red circles mark out the ideal Toe off and Heel Strike events.

during running. In this article, we focus only on the gait of walking. From Figure 2, we see that gait events such as Heel Strike (HS) and Toe Off (TO) of two feet play important roles in discriminating gait phases.

Ideally, a when foot is swinging, the pressure value at toe and heel should be zero since the foot is not touching the ground. Thus, the HS event and TO event can be easily identified as follows [National Institute of Neurological Disorders and Stroke 2016]:

—Given the preprocessed pressure time series *TS* of the left or right foot, if the sampling point s_i satisfies:
(1) the pressure value at the heel of s_{i+1}, $p_{H\_i+1} > 0$, and
(2) the pressure value at the heel of s_i, $p_{H\_i} = 0$,
then, s_i is an *HS* event.
—Given the preprocessed pressure time series *TS* of the left or right foot, if the sampling point s_i satisfies:
(1) the pressure value at the toe of $s_{i\text{-}1}$, $p_{T\_i\text{-}1} > 0$, and
(2) the pressure value at the toe of s_i, $p_{T\_i} = 0$,
then, s_i is a *TO* event.

However, due to hardware constraints and some special postures, when one foot is swinging, the real pressure readings at toes and heels do not always keep to zero. Figure 3 shows a sample of two-feet pressure readings of a subject. We can see that there are many irregular fluctuations in the swing phases. A worse fact is that the amplitude of such irregular fluctuations varies from subject to subject. If we adopt the intuitive idea directly, the irregular fluctuation will result in many pseudo HS and TO events.

To address the problem, we introduce a sliding window–based model to identify HS and TO events. According to Figure 3, we observe that the shape of reversed pressure readings at the toe is similar to that of the pressure readings at the heel and the points where TO events occur are similar to the points where HS events occur. We need to construct a model to identify only HS events since TO events will be identified by using the model to process the reversed pressure readings at the toe. For simplicity, we only present how to construct the model for identifying the HS events in this section. The model is described as follows.

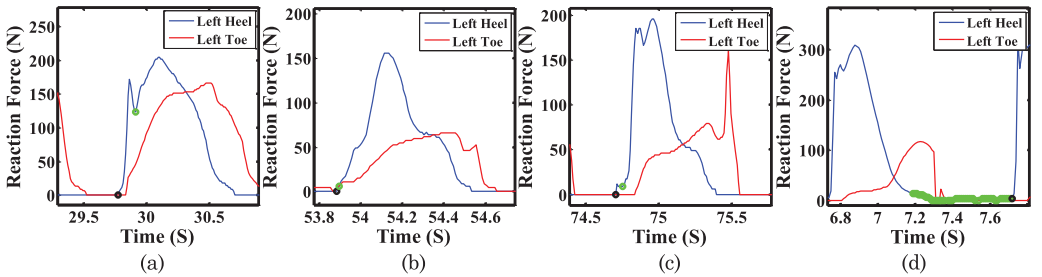

Fig. 4. A sample to illustrate that lack of any one constraint in the slide-window model will result in an error when identifying an HS event, where $l = 40, P = 15, M = 6$. (a) Without the first constraint, (b) without the second constraint, (c) without the third constraint, and (d) without the fourth constraint.

Given a pressure record s_i $(i = l, l + 1, \ldots, n)$, we construct a window as follows.

$$W(i, l) = [s_{i-l+1}, s_{i-l+2}, \ldots\ldots, s_i]$$

If s_i satisfies the following four constraints:

(1) $s_i.HeelValue < P$;
(2) $s_i.HeelValue \leq s_{i-1}.HeelValue$;
(3) Suppose M is the amount of the points satisfying

$$s_j.HeelValue \leq s_i.HeelValue, \text{ where } s_j \in W, \text{ then,}$$
$$M > AmountThreshold$$

(4) s_t does not satisfy conditions (1)∼(3), where

$$s_t \in \begin{cases} W(i + l - 1, l - 1), & i + l - 1 \leq n \\ W(n, n - i), & \text{others} \end{cases}$$

Then, s_i is a point where an HS event occurs.

Generally, the length of the slide window, l, should be set much smaller than the length of all the gait cycles, as shown in Figures 2 and 3. (Generally, the durations of all gait cycles are larger than 0.6s. Suppose that the sampling rate is 100Hz; the lengths of all gait cycles is larger than 60.) If l is too large, the window may contain more than one HS event. However, due to the fourth constraint, the window can only identify one HS event at a time and the other HS events will be missed. Therefore, we suggest setting l as 20∼30.

The first constraint is used to quickly filter out the points where HS events do not occur (shown as the green point in Figure 4(a)). In order to adapt to irregular fluctuation during swing phases, P should be set larger than the amplitude of the irregular fluctuation. Otherwise, an HS event may be filtered out falsely. Generally, we suggest setting P as 13∼20.

The second constraint is used to ensure that the extracted point is exactly the beginning of a pressure rising region (shown as the black point in Figure 4(b)). Without this constraint, the extracted point will have some extra time delay (shown as the green point in Figure 4(b)).

The third constraint enables the model to rule out the pseudo HS event caused by the nonmonotonic pressure changes (shown as the green point in Figure 4(c)).

In a gait cycle, there might be more than one point satisfying constraints 1∼3 (as shown in Figure 4(d)). However, in fact, one gait cycle contains only one HS event. If s_i is regarded as an HS event, it is impossible that the gait cycle, which starts from s_i,

contains another HS event. Before we identify all the gait events, we cannot know the length of the gait cycle, but we can infer that:

- If $i + l - 1 > n$, the window $W(i + l - 1, l)$, which starts with s_i, would not contain another HS event since the length of $W(i + l - 1, l)$, l, is much smaller than that of a gait cycle.
- Otherwise, the window $W(n, n - i + 1)$, which also starts with s_i, would not contain another HS event since the length of $W(n, n - i + 1)$, $n - i + 1$, is much smaller than that of a gait cycle.

By employing the proposed sliding window–based model, we can get all HS events of two feet from the pressure data of the left foot, TS_l, and the pressure data of the right foot, TS_r. According to Figure 3, we can see that the shape of reversed pressure readings at the toe is similar to that of the pressure readings at the heel, and the points where TO events occur are similar to the points where HS events occur. We can get all TO events of two feet by performing the proposed model on the reversed TS_l and TS_r.

Based on the extracted HS and TO events of two feet, the four gait phases—swing, initial double-support, single double-support and end double-support—can be defined as follows.

Definition 1. *Swing* (*Sw*) phase: It starts from the *TO* event of the objective foot and ends at the next *HS* event of the same foot.

Definition 2. *Initial double-support* (*Ids*) phase: It starts from the *HS* event of the objective foot and ends at the next *TO* event of the other foot.

Definition 3. *Single support* (*Ss*) phase: It starts from the *TO* event of the objective foot and ends at the next *HS* event of the other foot.

Definition 4. *End double-support* (*Eds*) phase: it starts from the *HS* event of the objective foot and ends at the next *TO* event of the other foot.

Due to the phase offset, the *Ids* phase of the objective foot is exactly the *Eds* phase of the other foot, and the *Ss* phase of the objective foot is exactly the *Sw* phase of the other foot. Figure 3 shows a sample of a two-feet pressure time series of an individual, in which the key gait events and gait phases are marked as well.

4.3. Gait-Feature Extraction and Selection

In this section, we present the features that we use to quantitatively characterize gaits. In order to measure gaits more comprehensively, we extract features from all three aspects of movement functions—stability, symmetry, and harmony—which are important metrics to assess movement function [Iosa et al. 2014].

To simplify the description, before computing the gait attributes, we calculate the duration of gait phase and transform each kind of gait phase into a time series. Finally, we can obtain a time series set *TSs* that contains eight time series as follows:

$$TSs = \{SwT_L, SwT_R, IdsT_L, IdsT_R, SsT_L, SsT_R, EdsT_L, EdsT_R\}, \tag{1}$$

where $SwT_{L/R}$, $IdsT_{L/R}$, $SsT_{L/R}$ and $EdsT_{L/R}$ denote the duration time series of swing phase, initial double-support phase, single-support phase and end-double support phase of two feet, respectively.

4.3.1. Stability. Gait stability is an important indicator of health [Ihlen et al. 2012; Yang et al. 2009] and is one of the typical symptoms of PD. Based on the extracted gait phases, we can measure gait stability from both gait rhythm and pressure amplitude. We use two features to measure the stability of gait rhythm.

(a) The coefficient of variation of the duration of gait phases (CV_{rhythm})

$$CV_{rhythm} = \sum_{i=1}^{8} CV(TSs(i)), \tag{2}$$

where $TSs(i)$ is generated from Equation (1). In particular,

$$CV(X) = 50 \times SD(X)/mean(X) \tag{3}$$

(b) The standard deviation of detrended time series of gait phase durations (DSD_{rhythm}). DSD_{rhythm} is formulated as

$$DSD_{rhythm} = \sum_{i=1}^{8} SD(Detrend(TSs(i))), \tag{4}$$

where $TSs(i)$ denotes the i_{th} time series of TSs, which is represented in Equation (1). $Detrend()$ denotes a detrending function, which is performed by taking the first difference of the time series. $SD()$ denotes the standard deviation.

In addition to stability of gait rhythm, we also investigate the stability of pressure amplitude. In this article, we employ two features to measure the stability of pressure amplitude.

(a) The coefficient of variation of maximum pressure in the standing phase (as shown in Figure 2, the *standing* phase is the ordered combination of the initial double-support phase, single-support phase and end double-support phase; $CV_{pressure}$).

Suppose that the time series of maximum pressure value in the standing phases of two feet are denoted as $PressureT_L$ and $PressureT_R$, respectively; then, $CV_{pressure}$ can be calculated as

$$CV_{pressure} = CV(PressureT_L) + CV(PressureT_L), \tag{5}$$

where $CV(X)$ is computed by Equation (3).

(b) The standard deviation of detrended time series of maximum pressure in the standing phase ($DSD_{pressure}$). $DSD_{pressure}$ can be computed as

$$DSD_{pressure} = SD(Detrend(PressureT_L)) + SD(Detrend(PressureT_R)) \tag{6}$$

where the functions $Detrend()$ and $SD()$ have been introduced in Equation (4).

On the one hand, CV_{rhythm} and $CV_{pressure}$ indicate the discreteness of the duration of all kinds of gait phases and the discreteness of the maximum pressure value during the standing phases, respectively. The advantage of CV_{rhythm} and $CV_{pressure}$ is that they can minimize the effect of individual differences in walking speed and weight, respectively. Generally, more stable gait results in smaller CV_{rhythm} and $CV_{pressure}$. On the other hand, DSD_{rhythm} and $DSD_{pressure}$ indicate the relative fluctuation of duration of all kinds of gait phases and the relative fluctuation of maximum pressure values during the standing phases, respectively. The advantage of DSD_{rhythm} and $DSD_{pressure}$ is that they minimize the effect of changes between successive phase elements. Generally, a more stable gait results in smaller DSD_{rhythm} and $DSD_{pressure}$.

4.3.2. Symmetry. There is evidence proving that movement asymmetry is commonly observed along with a decline in health status [Yogev et al. 2007]. Particularly, PD patients may present very asymmetrical gaits [Baltadjieva et al. 2006]. Similar to gait stability, we measure gait symmetry from both gait rhythm and pressure amplitude.

Given the time series of the swing phase of two feet, SwT_L and SwT_R (as shown in Figure 5), we can see that P_{L_i} and P_{R_i} ($i = 1, 2, \ldots, m$) occur alternately over time. Without loss of generality, suppose that P_L occurs first (i.e., one first step forward with

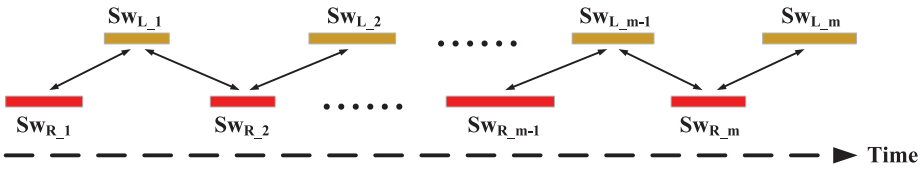

Fig. 5. The process of computing gait symmetry of gait rhythm.

the left foot), we can measure the total duration difference, *TotalDiff*, between two feet as follows:

$$TotalDiff(SwT_L, SwT_R) = |P_{L\_m} - P_{R\_m}| + \sum_{i=1}^{m-1} (|P_{L\_i} - P_{R\_i}| + |P_{L\_i} - P_{R\_i+1}|). \quad (7)$$

Specifically, the duration difference between two elements deriving from two feet, respectively, are taken into consideration only if the two elements are adjacent over time. Obviously, more elements will result in larger *TotalDiff*. To eliminate the effect of different amounts of elements, the average difference, *AvgDiff*, is computed as

$$AvgDiff(SwT_L, SwT_R) = \frac{TotalDiff(SwT_L, SwT_R)}{2m-1}. \quad (8)$$

Afterwards, to minimize the effect of difference of absolute value between different individuals, $AvgDiff\ (SwT_L, SwT_R)$ is normalized by the mean value of SwT_L and SwT_R. Thus, the symmetry of the swing phase is finally computed as

$$Sym(SwT_L, SwT_R) = \frac{2AvgDiff(SwT_L, SwT_R)}{mean(SwT_L) + mean(SwT_R)}. \quad (9)$$

Similarly, we can obtain $Sym(IdsT_L, IdsT_R)$, $Sym(SsT_L, SsT_R)$, and $Sym(EdsT_L, EdsT_R)$. Finally, the symmetry of gait rhythm, Sym_{rhythm}, is represented as the sum of four kinds of gait-phase symmetry as follows.

$$\begin{aligned} Sym_{rhythm} = {} & Sym(SwT_L, SwT_R) + Sym(IdsT_L, IdsT_R) \\ & + Sym(SsT_L, SsT_R) + Sym(EdsT_L, EdsT_R). \end{aligned} \quad (10)$$

In addition, we investigate the symmetry of pressure amplitude. Suppose that the time series of maximum pressure value in the standing phases of two feet are denoted as $PressureTS_L$ and $PressureTS_R$, respectively. Then, the symmetry of pressure amplitude, $Symm_{pressure}$, can be calculated as

$$Sym_{pressure} = Sym(PressureT_L, PressureT_R), \quad (11)$$

where the function $Sym()$ is defined by Equation (9).

Sym_{rhythm} indicates the average duration difference of four kinds of gait phases between two feet. $Sym_{pressure}$ indicates the average difference of maximum pressure value during the standing phase between two feet (as the pressure value keeps to 0 during the swing phase, only the standing phase is considered in $Sym_{pressure}$). In general, poor gait symmetry would accompany larger Sym_{rhythm} and Sym_{rhythm}.

4.3.3. Harmony. Harmony is one of the intrinsic features of human gait, which contributes to efficient and smooth movements during walking [Iosa et al. 2012], and facilitates repetitive walking in the healthy individual [Iosa et al. 2013]. Intuitively, gait harmony refers to the proportions between different gait phases. For example, during one gait cycle, the proportion between the standing phase and the swing phase should be about 60% versus 40%, while significant alterations of these proportions are

usually identified as signs of pathological gait [Winter et al. 1990]. In this article, we employ two proportions to describe gait harmony:

(a) The average proportion of the swing phase in a gait cycle (P_1);

$$P_1 = \frac{1}{2}(P_{1L} + P_{1R}), \tag{12}$$

where P_{1L} can be calculated based on Equation (13):

$$P_{1L} = \frac{1}{m}\sum_{i=1}^{m} \frac{SwT_L(i)}{SwT_L(i) + IdsT_L(i) + SsT_L(i) + EdsT_L(i)}, \tag{13}$$

where SwT_L, $IdsT_L$, SsT_L, and $EdsT_L$ denote the time series of four kinds of gait phases of the left foot, respectively (as mentioned in Equation (1)). Similarly, we can compute P_{1R}.

(b) The average proportion of the single support phase in the standing phase (P_2):

$$P_2 = \frac{1}{2}(P_{2L} + P_{2R}), \tag{14}$$

where P_{2L} can be calculated based on Equation (15):

$$P_{2L} = \frac{1}{m}\sum_{i=1}^{m} \frac{SsT_L(i)}{IdsT_L(i) + SsT_L(i) + EdsT_L(i)}. \tag{15}$$

Similarly, we can compute P_{2R}.

In order to gain better gait pattern classification performance and reduce the computational complexity in these successive steps, we retain only the features that contribute significantly to classifying gait patterns. In this article, we use Information Gain (IG) [Kullback 1959] to measure the effectiveness of candidate features. Specifically, features whose IG is smaller than 0.1 will be discarded. The remaining features will be used as input for the gait pattern classifier, which are organized as a feature vector.

4.4. Gait Pattern Classification

In our previous work [Wang et al. 2015], we used a BP neural network to construct the classification model and achieved good recognition accuracy. To further improve the performance, in this article, we use Adaboost [Jiawei and Kamber 2006], a popular ensemble method to construct a hybrid model, which combines several BP neural network models as basic models.

The basic idea of Adaboost [Han et al. 2006] can be described as follows. Several BP neural network classifiers are learned iteratively. The next classifiers will pay more attention to the training tuples incorrectly classified in the previous classifiers. Finally, the votes of each classifier are combined with different weights, which are positively related to the accuracy of different classifiers.

Formally, given a dataset D, containing d-labeled training tuples, ($\boldsymbol{X}_1$, y_1), ($\boldsymbol{X}_2$, y_2), ..., ($\boldsymbol{X}_d$, y_d), where y_i is the label of $\boldsymbol{X}_i$, Adaboost assigns an equal weight $1/d$ to each training tuple. Then, it trains k BP neural network classifier models iteratively. For the j_{th} ($j = 1, 2, \ldots, k$) model M_j, first, a repeatable sampling method is used to generate a training dataset D_j from D. The size of D_j is equal to that of D. The chance of each tuple being selected depends on its weight. The ones with higher weights are more likely to be selected. Second, M_j is learned on D_j, and its error $error(M_j)$ is computed using D_j as a testing dataset:

$$error(M_j) = \sum_{t=1}^{d} w_t \times err(\boldsymbol{X}_t), \tag{16}$$

Table I. Statistics of Subjects

	Parkinson's disease patients	Healthy individuals
Total number	93	72
Percentage of men	63%	55%
Age	66.301 ± 9.500	63.658 ± 8.641
Height (m)	1.674 ± 0.086	1.682 ± 0.085
Weight (kg)	72.396 ± 11.960	72.764 ± 12.347
UPDRS	31.615 ± 11.825	0.477 ± 0.876
UPDRSm	19.319 ± 7.656	0.477 ± 0.876
H&Y	2.258 ± 0.343	0 ± 0

Age, height, weight, UPDRS, UPDRSm, and H&Y are shown as: mean $\pm$ std.

where $\boldsymbol{X}_t \in D_j$, w_t is the weight of $\boldsymbol{X}_t$. $err(\boldsymbol{X}_t)$ is the misclassification error of $\boldsymbol{X}_t$. If $\boldsymbol{X}_t$ is misclassified, $err(\boldsymbol{X}_t) = 1$; otherwise, $err(\boldsymbol{X}_t) = 0$. If the performance of M_j is so poor ($error(M_j) > 0.5$), then it is abandoned and a new M_j will be generated from a new D_j. Third, the weights of all the training tuples are updated. Correctly classified tuples will get higher weights; incorrectly classified tuples will get lower weights. Specifically, if $\boldsymbol{X}_t$ is correctly classified, then

$$w_t = w_t \times \frac{error(M_j)}{1 - error(M_j)}. \tag{17}$$

The weights of the tuples reflect the difficulty of correctly classifying them. Higher weight indicates that they are more often misclassified. The weights of tuples will be used to generate the training dataset D_{j+1} for the next model M_{j+1}.

Given a tuple $\boldsymbol{X}$, Adaboost uses the weighted combination of BP neural network models to predict its label. Specifically, the model with smaller error (i.e., more accurate), will be assigned with a higher weight. The weight of model M_j can be calculated as follows:

$$weight(M_j) = \log \frac{1 - error(M_j)}{error(M_j)}. \tag{18}$$

Then, for each class c, the weights of the models that classify $\boldsymbol{X}$ into c will be accumulated. Finally, the one with the largest weight sum will be regarded as the class for $\boldsymbol{X}$.

5. EXPERIMENT EVALUATION

In this section, we report the evaluation results of the proposed framework.

5.1. Experimental Setup

The datasets used in this work are derived from the study by Hausdorff et al. [2007], Yogev et al. [2005], and Frenkel-Toledo et al. [2005], which consist of gait data from 93 patients with idiopathic PD and 72 healthy individuals. The PD patients were evaluated by a movement-disorder neurologist according to the unified Parkinson's disease rating scale (UPDRS), UPDRS motor (UPDRSm), and Hoehn and the Yahr (H&Y) scale. The detailed statistics of the subjects are shown in Table I. The database includes the vertical ground reaction force records of subjects as they walked under two different states (walking normally and working on serial-7 subtraction) at a self-selected pace for approximately 2min on level ground. In this study, we used only the data from subjects walking normally. Underneath each foot, there were 8 pressure sensors that measure force (in Newtons) as a function of time. The sampling frequency of these 16 sensors is 100Hz. Actually, different heel height and hardness of the shoes may slightly change the gait kinematics [Cowley et al. 2009; Bendix et al. 1984]. For

Table II. Summary of the Features Extracted by the Proposed Method

	PT	CO	*P* value	*IG*
$CV_{pressure}$	48.81 ± 17.80	37.33 ± 12.70	1.97E-05	0.105
$DSD_{pressure}$	107.53 ± 32.12	108.33 ± 39.99	0.89	0
CV_{rhythm}	50.05 ±16.34	36.41 ± 8.96	5.97E-09	0.164
DSD_{rhythm}	0.16 ± 0.09	0.11 ± 0.03	3.47E-07	0.186
$Sym_{pressure}$	0.88 ± 0.29	0.74 ± 0.23	8.22E-04	0
Sym_{rhythm}	0.11 ± 0.07	0.06 ± 0.02	1.56E-09	0.327
P_1	0.34 ± 0.03	0.36 ± 0.01	1.08E-07	0.143
P_2	0.52 ± 0.06	0.56 ± 0.03	6.43E-08	0.143

These features are shown as: mean ± std.
P value derived from a two-sample t-test.
H_0: features from PT and Control are consistent with one single distribution.

example, high-heeled shoes will slightly lengthen the stand phase and end double-support phase while shortening the stride interval and single-support phase, resulting in smaller P_1 and P_2. In addition, soft shoes may mildly magnify the deviation of the stride intervals, resulting in a larger CV_{rhythm}. However, the plantar dataset that we use was collected from the subjects who were tested using the same system (a pair of shoes embedded with pressure sensors and a recording unit carried on the waist) [Hausdorff et al. 2007]. Hence, the negative impacts caused by the shoes basically can be ignored.

We used four baseline methods: the method in Wu and Krishnan [2010], our previous work [Wang et al. 2015], the method in Sant'Anna et al. [2011] and the method using the same features as the proposed method but recognizing gait patterns by employing a single BP neural network. The first method [Wu and Krishnan 2010] extracted the stride intervals (i.e., stride-to-stride duration) standard deviation (σ) and STC, and employs the LS-SVM to classify gait patterns corresponding to healthy individuals and PD patients. The second baseline method, in our previous work [Wang et al. 2015], is similar to the proposed framework, but it used a simple method to identify gait events HS and TO, which led to time offset in most of the extracted gait events. These gait events with time offset will finally affect the subsequent quantitative analysis. In addition, the classification model was constructed by a neural network, while we constructed a hybrid classification model by using the Adaboost algorithm in this article to achieve high recognition accuracy. The basic idea of the third baseline method [Sant'Anna et al. 2011] was to classify gait patterns based on gait symmetry, which was computed using a symbolic processing method. To further illustrate the advantage of the proposed framework, we replaced Adaboost (the method used in this article to construct the gait pattern recognition model) with a BP neural network, that is, the fourth baseline method, which used the same gait features.

We used two evaluation metrics: AUC [Bradley 1997] and a pair of Precision and Recall. The formal definition of AUC and the pair (Precision, Recall) can be found in our previous work [Wang et al. 2015]. Generally speaking, high AUC, and Precision and Recall, indicate good gait pattern classification performance.

Our evaluations were conducted in MATLAB on an Intel (R) Xeon (R) CPU E5-2620 with 32GB RAM running Windows 8.

5.2. Evaluation Results

5.2.1. Gait Feature Comparison. A summary of the features extracted in this study are shown in Table II. Specifically, we can obtain the following three observations according to Table II.

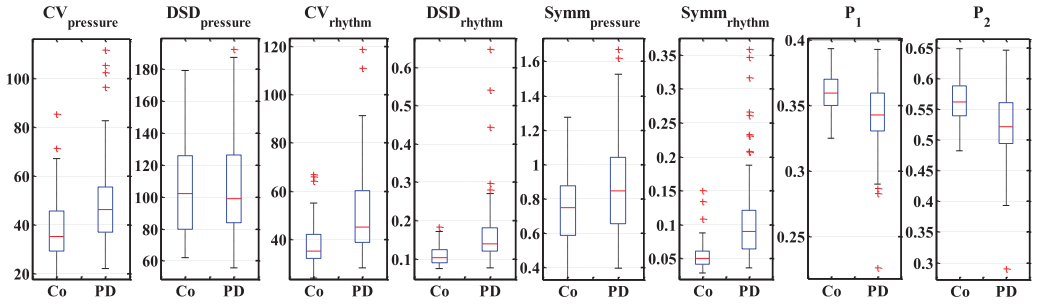

Fig. 6. The box-plots of all features extracted by the proposed method. The whiskers represent the smallest and largest observations, the edges of the box correspond to the lower and upper quartiles, the horizontal line indicates the median, and the plus sign marks probable outliers. Co: control, PD: PD patients.

—Except for $DSD_{pressure}$, both the mean value and standard deviation of the features measuring the gait stability of healthy individuals (i.e., $CV_{pressure}$, CV_{rhythm}, $DSD_{rhythm\,e}$) are obviously smaller than that of PD patients. It indicates that the gait stability of PD patients is weaker than that of healthy individuals. Moreover, *the P* value of the features between healthy individuals and PD patients is quite small, that is, the three features of PD patients and healthy individuals do not follow the same distribution, which is helpful for gait-pattern classification.
—Similarly, the mean value and standard deviation of the features measuring the gait symmetry of healthy individuals (i.e., $Symmetry_{pressure}$ and $Symmetry_{rhythm}$) are smaller than that of PD patients. Especially in the case of $Symmetry_{rhythm}$, the mean value and standard deviation of healthy individuals are much smaller than that of PD patients. It indicates that the gait-rhythm symmetry of PD patients is significantly weaker than that of healthy subjects, which is consistent with common sense. In addition, the P value of $Symmetry_{rhythm}$ between healthy individuals and PD patients is smaller than 1.56E-09, which is smaller enough to indicate that the $Symmetry_{rhythm}$ measures of PD patients and healthy individuals do not follow the same distribution.
—The features measuring gait harmony (i.e., P_1 and P_2) of PD patients are both smaller than that of healthy individuals. In other words, the standing phase proportion of PD patients is generally greater than that of healthy individuals, and the double-support phase of PD patients is generally greater than that of healthy individuals. The reason is that decreased ability to maintain balance forces PD patients to lengthen the standing phase, especially the double-support phase, to keep balance. Furthermore, the P value of P_1 and P_2 is small enough to demonstrate that these two features of PD patients and healthy individuals do not follow the same distribution.

The visualization of the extracted features is shown in Figure 6. We can see that, except for $DSD_{pressure}$ and $Symmetry_{pressure}$, the distribution of the features between healthy individuals and PD patients is quite different, which corroborates the observations obtained from Table II. However, the overlap of the distribution is so large compared to the difference in median that we cannot distinguish gait patterns by any single feature. Therefore, we organized all the features as a vector and used it as the input of the classifier. To optimize the input vector, in this article, we used IG to select features. If the IG value of one feature is smaller than 0.1, it will be discarded. The IG of extracted features is presented in Table II. Consistent with the observation from Figure 6, $DSD_{pressure}$ and $Symmetry_{pressure}$ contribute less to gait pattern classification. Therefore, these two features were discarded, and the final input vector was chosen as follows:

$$V_{input} = \langle CV_{pressure}, CV_{rhythm}, DSD_{rhythm}, Symmetry_{rhythm}, P_1, P_2 \rangle .$$

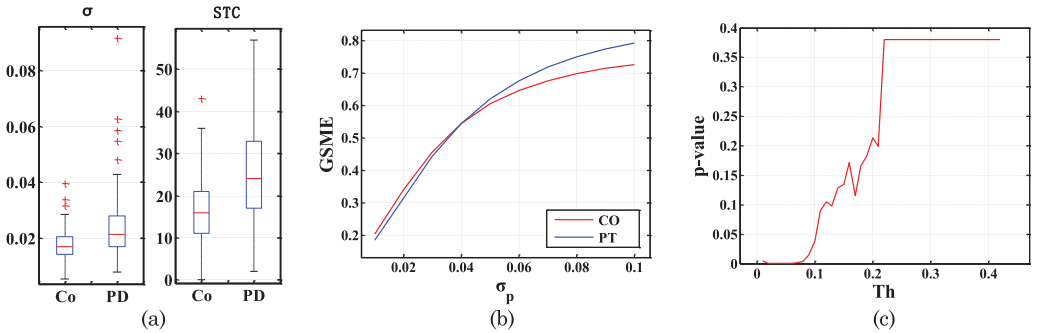

Fig. 7. Optimal features extracted by the first baseline method [Wu and Krishnan 2010]. (a) the box-plots of the standard deviation of gait rhythm ($\sigma_{\mathrm{P}} = 0.01$) and Signal Turns Count, STC ($Th = 0.02$); (b) GMSE as σ_{P} is varied; (c) P value of STC as th is varied.

Table III. Summary of the Features Extracted by the First Baseline Method [Wu and Krishnan 2010]

Features	PT	CO	*P* value
σ ($\sigma_P = 0.01$)	0.018 ± 0.006	0.024 ± 0.012	1.16E-4
STC ($Th = 0.02$)	16.931 ± 8.794	25.108 ± 11.143	8.78E-4

These features are shown as: mean $\pm$ std.

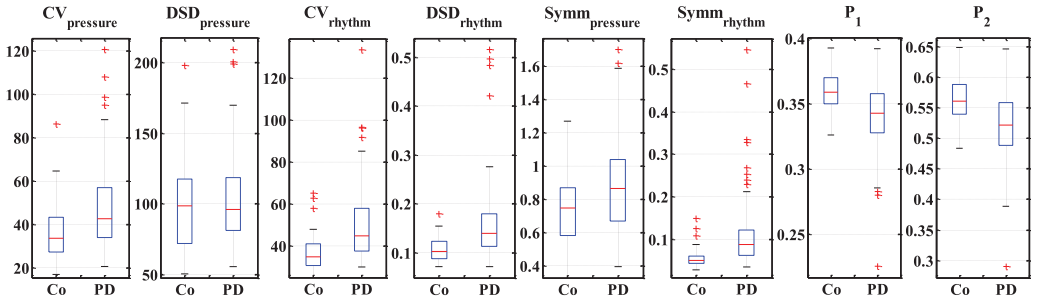

Fig. 8. The box plots of all features extracted by the second baseline method [Wang et al. 2015].

The features σ (standard deviation of gait rhythm) and STC extracted by the first baseline method are shown in Figure 7 (the detailed feature optimization process can be seen in our previous work [Wang et al. 2015]). A summary of σ and STC is shown in Table III. We can see that the P values of these two features are much greater than that of the features in V_{input}, indicating that the features extracted by the first baseline method are less effective for gait pattern recognition than the features in V_{input}.

The features extracted in our previous work [Wang et al. 2015] are shown in Figure 8. We can see that there is no obvious difference compared to the features extracted by the proposed method. A summary of these features is shown in Table IV. The P values of DSD_{rhythm}, P_1 and P_2, are slightly smaller than the P values of the corresponding features extracted based on the proposed method in this article. However, the P values of $CV_{pressure}$, CV_{rhythm}, and Sym_{rhythm} are larger than the P values of the corresponding features extracted in this article. It indicates that features DSD_{rhythm}, P_1, and P_2 extracted by the second baseline method are more effective for gait-pattern recognition, while the other three features $CV_{pressure}$, CV_{rhythm}, and Sym_{rhythm} extracted by the proposed method in this article are more useful.

Table IV. Summary of the Features Extracted by the Second Baseline Method [Wang et al. 2015]

Features	PT	CO	*P* value
$CV_{pressure}$	48.06 ± 20.49	36.20 ± 12.04	2.26E-05
$DSD_{pressure}$	103.28 ± 32.99	101.60 ± 37.96	0.7622
CV_{rhythm}	50.29 ± 17.52	36.64 ± 8.01	6.30E-09
DSD_{rhythm}	0.16 ± 0.08	0.11 ± 0.02	9.90E-08
$Symm_{pressure}$	0.89 ± 0.30	0.74 ± 0.23	5.10E-04
$Symm_{rhythm}$	0.11 ± 0.08	0.06 ± 0.02	7.77E-08
P_1	0.34 ± 0.03	0.36 ± 0.01	5.27E-08
P_2	0.52 ± 0.06	0.56 ± 0.03	3.21E-08

These features are shown as: mean ± std.
P value derived from a two-sample t-test.
H_0: features from PT and Control are consistent with one single distribution.

Table V. Summary of the Symmetry Extracted by the Third Baseline Method [Sant'Anna et al. 2011]

Features	PT	CO	*P* value
Symb8	22.23 ± 6.52	19.42 ± 3.72	0.001
Symb10	23.44 ± 6.28	20.37 ± 3.63	3.04E-04
Symb12	23.99 ± 5.93	20.95 ± 3.67	1.93E-04
Symb14	24.39 ± 5.71	21.08 ± 3.52	2.72E-05
Symb16	24.66 ± 5.69	21.31 ± 3.28	1.58E-05
Symb18	24.99 ± 5.66	21.59 ± 3.20	1.02E-05

These features are shown as: mean ± std.
P value derived from a two-sample t-test.
H_0: features from PT and Control are consistent with one single distribution.

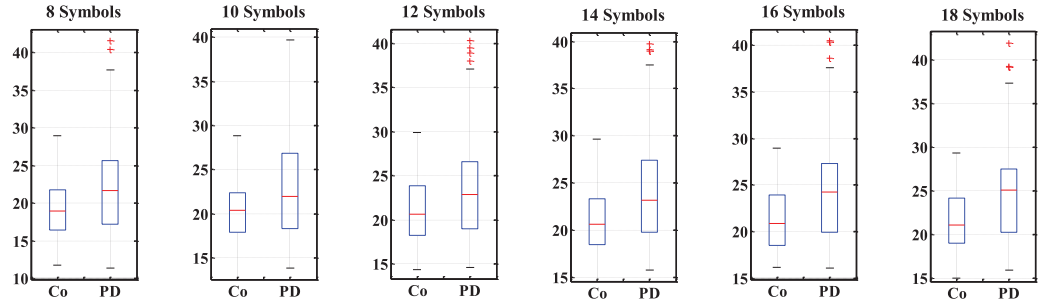

Fig. 9. The box plots of the features generated by the third baseline method [Sant'Anna et al. 2011]. The subfigures 8 Symbols ∼ 18 Symbols correspond to the symmetry generated under 8 ∼ 18 symbols, respectively.

Features extracted by the third baseline method [Sant'Anna et al. 2011] are shown in Figure 9. Different numbers of symbols will generate different symmetry measures. In this article, we computed symmetries by setting the amount of symbols as even numbers ranging from 8 to 18. We can see that, compared with the features in V_{input}, the distribution difference between healthy individuals and PD patients is not obvious. Moreover, there is no distinct change as the number of symbols varies. A summary of these features is presented in Table V. Comparing with the *P* value of the features in Table II, the *P* values of these features are larger overall. In other words, the features extracted based on the third baseline method are less effective than the features extracted by our proposed method for gait-pattern recognition.

5.2.2. Gait-Pattern Classification Results. In order to get robust verification results, 10 times 10-fold cross-validation [Kohavi et al. 1995] was applied to the proposed method and the four baseline methods. Specifically, for each 10-fold cross-validation, a new random seed was used to repartition the dataset into 10 subsets. Each subset was used as the test set in turn and the other 9 subsets were used as training sets. The classification results of the proposed method and baseline methods are shown in Table VI (note that, for the third baseline method, the symbol number was set as 18, where the *P* value is the smallest, as attributes with a smaller *P* value will generate better classification results).

By comparing the 10 times 10-fold cross-validation classification results in Table VI, we find that the proposed method is:

Table VI. Classification Result Comparison

First baseline method				Second baseline method				Third baseline method			
Num.	**Pre.**	**Rec.**	**AUC**	**Num.**	**Pre.**	**Rec.**	**AUC**	**Num.**	**Pre.**	**Rec.**	**AUC**
1	0.627	0.584	0.686	**1**	0.846	0.828	0.866	**1**	0.675	0.675	0.602
2	0.627	0.584	0.687	**2**	0.813	0.839	0.862	**2**	0.634	0.624	0.677
3	0.627	0.584	0.687	**3**	0.835	0.817	0.837	**3**	0.650	0.643	0.677
4	0.627	0.584	0.686	**4**	0.823	0.849	0.862	**4**	0.634	0.636	0.602
5	0.627	0.584	0.686	**5**	0.828	0.828	0.846	**5**	0.625	0.655	0.591
6	0.626	0.584	0.686	**6**	0.859	0.849	0.844	**6**	0.628	0.627	0.559
7	0.627	0.584	0.686	**7**	0.821	0.839	0.856	**7**	0.664	0.670	0.656
8	0.627	0.584	0.686	**8**	0.809	0.817	0.851	**8**	0.660	0.690	0.645
9	0.627	0.584	0.687	**9**	0.854	0.817	0.847	**9**	0.599	0.596	0.602
10	0.627	0.584	0.687	**10**	0.828	0.828	0.842	**10**	0.650	0.633	0.667
Avg	**0.627**	**0.584**	**0.686**	**Avg**	**0.831**	**0.831**	**0.851**	**Avg**	**0.642**	**0.645**	**0.628**
Fourth baseline method				**Our proposed method**							
Num.	**Pre.**	**Rec.**	**AUC**	**Num.**	**Pre.**	**Rec.**	**AUC**				
1	0.828	0.828	0.858	**1**	0.883	0.892	0.889				
2	0.837	0.828	0.861	**2**	0.912	0.892	0.896				
3	0.846	0.828	0.866	**3**	0.863	0.882	0.908				
4	0.830	0.839	0.856	**4**	0.871	0.871	0.882				
5	0.837	0.828	0.870	**5**	0.853	0.871	0.901				
6	0.837	0.828	0.866	**6**	0.879	0.860	0.889				
7	0.821	0.839	0.856	**7**	0.879	0.860	0.894				
8	0.837	0.828	0.868	**8**	0.902	0.892	0.894				
9	0.828	0.828	0.864	**9**	0.878	0.849	0.885				
10	0.839	0.839	0.865	**10**	0.871	0.871	0.895				
Avg	**0.834**	**0.831**	**0.863**	**Avg**	**0.879**	**0.874**	**0.893**				

Results keep three decimal places.

—significantly superior to the first baseline method by 39.4%, on average, in terms of precision; 30.2%, on average, in terms of recall; and 40.2%, on average, in terms of AUC.
—slightly superior to the second baseline method by 5.2%, on average, in terms of precision; 4.9%, on average, in terms of recall; and 5.8%, on average, in terms of AUC
—significantly superior to the third baseline method by 35.5%, on average, in terms of precision; 42.2%, on average, in terms of recall; and 36.9%, on average, in terms of AUC
—slightly superior to the fourth baseline method by 5.2%, on average, in terms of precision; 3.5%, on average, in terms of recall; and 5.4%, on average, in terms of AUC

The ROC curves of the proposed method and four baseline methods are shown in Figure 10, which indicates that our method is able to achieve a high detection rate while keeping a low false-alarm rate. The proposed method, in the worst case, is able to accurately recognize over 85% of PD patients while keeping the false-alarm rate lower than 16%. Specifically, for the 8th 10-fold cross-validation (E-8), over 90% of PD patients can be accurately recognized with a false-alarm rate of less than 13%. In contrast, for the four baseline methods, in the best case, to achieve a detection rate over 85%, they have to suffer over 68%, 22%, 67%, and 25% false-alarm rates, respectively. Generally speaking, the AUC values of all four baseline methods are smaller than that of the proposed method.

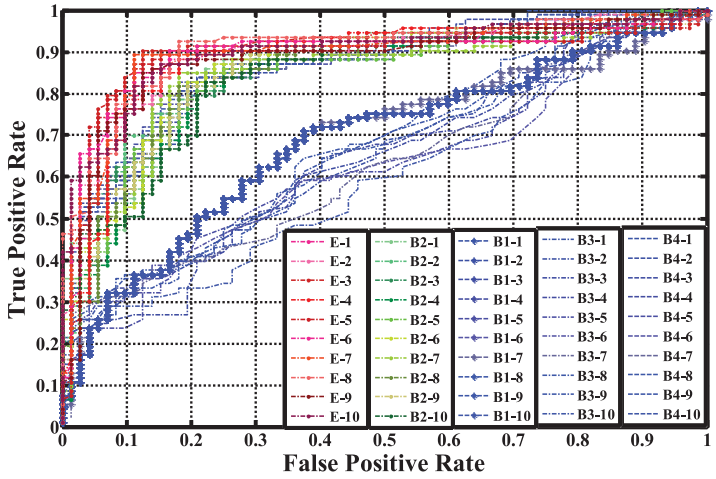

Fig. 10. The ROC curves of the proposed method and baseline methods. E-1 ∼ E-10, the ROC curves of 10 times 10-fold cross-validation of our proposed method. B1-1 ∼ B1-10, the ROC curves of 10 times 10-fold cross-validation of the first baseline method. B2-1 ∼ B2-10, the ROC curves of 10 times 10-fold cross-validation of the second baseline method. B3-1 ∼ B3-10, the ROC curves of the third baseline method as the symbol number is set to the even number from 8 to 18. B4-1 ∼B4-10, the ROC curves of 10 times 10-fold cross-validation of the fourth baseline method.

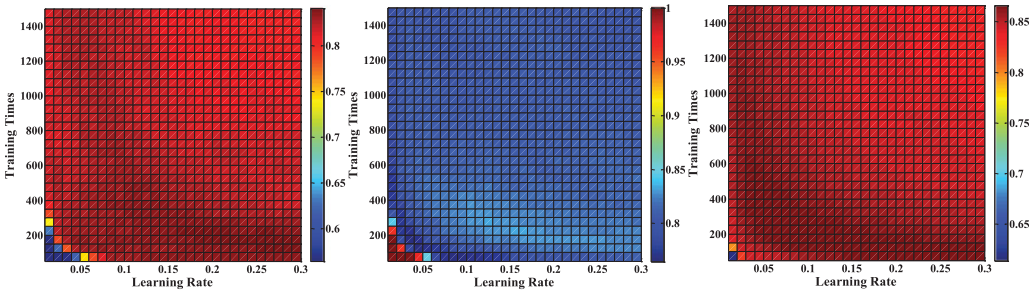

Fig. 11. Average Precision, Recall and AUC of BP neural network models as training times and learning rate are varied. (a) Average Precision of 10 times 10-fold cross-validation on the dataset as training times and learning rate are varied. (b) Average Recall of 10 times 10-fold cross-validation on the dataset as training times and learning rate are varied. (c) Average AUC of 10 times 10-fold cross-validation on the dataset as training times and learning rate are varied.

5.2.3. Parameter Selection. In the gait-pattern classification method, there are three key parameters: the number of BP neural network models (as the basic classification models), the learning rate, and the training time, which directly affect the performance of gait-pattern classification. To achieve high recognition accuracy, we first selected the appropriate learning rate and training time for all BP neural network models to obtain the optimized basic classification models. Figure 11 shows the average precision, recall, and AUC of BP neural network models, as the training time and learning rate vary. According to Figure 11(a), we can see that when the training time is larger than 350 and the learning rate is larger than 0.05, the precision tends to be steady at about 0.83. According to Figure 11(b), we observe that the recall tends to be steady at about 0.825 when the training time is larger than 300 and the learning rate is larger than 0.05. Based on Figure 11(c), we can see that the AUC value tends to be steady at about 0.86

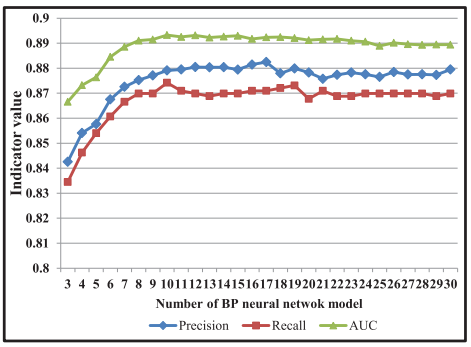

Fig. 12. Average Precision, Recall and AUC of gait-pattern classification method as the number of BP neural network varies.

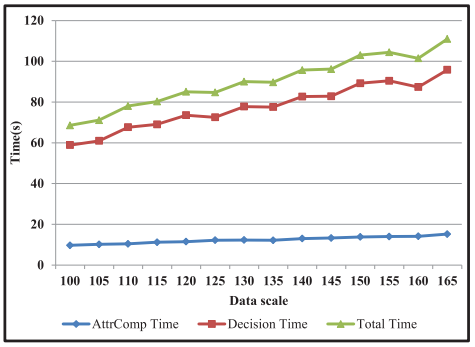

Fig. 13. Processing time on different data scales, in which number of BP neural network and training time are fixed to 10 and 600,respectively

when the training time is larger than 250 and the learning rate is larger than 0.04. In this article, we suggest that setting training times = 600 and learning rate = 0.06 for each BP neural network is a fair trade-off for the proposed method on the used dataset. Afterwards, we selected an appropriate number of BP neural network models to construct the final hybrid classifier. According to Figure 12, we observe that the best performance is achieved when the number of BP neural network models is set to 10.

5.2.4. Computation Complexity Analysis. Figure 13 demonstrates how the processing time changes with respect to the data scale, in which the number of the BP neural network models and the training time was set to 10 and 600, respectively. We observe that both the time consumed to extract features (AttrComp Time) and the time consumed to classify gait patterns (Decision Time) increase approximately linearly as the data scale increases. In other words, the total time complexity increases linearly as the data scale increases, indicating that the computation time of our method will not increase dramatically as the data scale grows.

6. CONCLUSION AND FUTURE WORK

In this article, we investigated the problem of extracting fine-grained gait features from plantar pressure data and identifying gait patterns in PD patients. Our work was motivated by the fact that fine-grained gait features are much more efficient in revealing movement function and discovering PD gait patterns. To address the problem, we first discriminated four gait phases (i.e., swing phase, initial double-standing phase, single-standing phase, end double-standing phase) from plantar pressure data during walking by using a sliding window–based model. Based on the gait phases of two feet, we further extracted and selected a set of features to quantitatively describe the gait from three aspects of movement functions: stability, symmetry and harmony. Third, we classified gait patterns corresponding to PD patients and healthy individuals based on a hybrid classification model, which contained several BP neural network classification models. Finally, we evaluated the proposed framework on an open dataset, which includes the real plantar pressure data of 93 idiopathic PD patients and 72 healthy individuals. The results demonstrate that our framework outperforms four baseline methods.

In the future, we plan to extend our work in two directions. First, we will attempt to use other types of data, such as acceleration and angular velocity, to further improve gait-recognition performance. Second, we would like to apply our approach to address other motor nervous system diseases, for example, stroke and Huntington's chorea.

ACKNOWLEDGMENTS

The authors would like to express their special appreciation to W. Zhao, Y. Song, J. Deng and F. Liu, for their assistance in the experiments.

REFERENCES

R. Baltadjieva, N. Giladi, L. Gruendlinger, C. Peretz, and J. M. Hausdorff. 2006. Marked alterations in the gait timing and rhythmicity of patients with de novo Parkinson's disease. *European Journal of Neuroscience* 24, 6, 1815–1820.

T. Bendix, S. S. Steen, and K. Klausen. 1984. Lumbar curve, trunk muscles, and line of gravity with different heel heights. *Spine* 9, 2, 223.

A. P. Bradley. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30, 7, 1145–1159.

E. E. Cowley, T. L. Chevalier, and N. Chockalingam. 2009. The effect of heel height on gait and posture. *Journal of the American Podiatric Medical Association* 99, 6, 512–518.

S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, and J. M. Hausdorff. 2005. Effect of gait speed on gait rhythmicity in Parkinson's disease: Variability of stride time and swing time respond differently. *Journal of Neuro Engineering and Rehabilitation* 2, 3.

Z. Gandomkar and F. Bahrami. 2012. Proposing a new set of features based on frieze pattern to discriminate normal and abnormal gait. *International Conference on Biomedical Engineering* 354–359.

B. M. Gupta and A. Bala. 2013. Parkinson's disease in India: An analysis of publications output during 2002–2011. *International Journal of Nutrition, Pharmacology, Neurological Diseases* 3, 3, 254.

J. Han, H. S. Jeon, W. J. Yi, B. S. Jeon, and K. S. Park. 2009. Adaptive windowing for gait phase discrimination in parkinsonian gait using 3-axis acceleration signals. *Medical & Biological Engineering & Computing* 47, 11, 1155–1164.

M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, A. L. Peterson, D. Kaliton, and A. L. Goldberger. 2000. Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. *Journal of Applied Physiology* 88, 6, 2045–2053.

J. M. Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger, C. Peretz, and N. Giladi. 2007. Rhythmic auditory stimulation modulates gait variability in Parkinson's disease. *European Journal of Neuroscience* 26, 8, 2369–2375.

E. A. Ihlen, T. Goihl, P. B. Wik, O. Sletvold, J. Helbostad, and B. Vereijken. 2012. Phase-dependent changes in local dynamic stability of human gait. *Journal of Biomechanics* 45, 13, 2208–2214.

M. Iosa, A. Fusco, F. Marchetti, G. Morone, C. Caltagirone, S. Paolucci, and A. Peppe. 2013. The golden ratio of gait harmony: Repetitive proportions of repetitive gait phases. *BioMed Research International* 2013, Article ID918642, 7 pages.

M. Iosa, F. Paradisi, S. Brunelli, and A. Delussu, etc. 2014. Assessment of gait stability, harmony, and symmetry in subjects with lower-limb amputation evaluated by trunk accelerations. *JRRD* 51, 4, 623–634.

Marco Iosa, Francesco Paradisi, and Stefano Brunelli, et al. 2014. Assessment of gait stability, harmony, and symmetry in subjects with lower-limb amputation evaluated by trunk accelerations. *Journal of Rehabilitation Research and Development* 51, 4, 623–634.

M. Iosa, T. Marro, S. Paolucci, and D. Morelli. 2012. Stability and harmony of gait in children with cerebral palsy. *Research in Developmental Disabilities* 33, 1, 129–135.

Han Jiawei and Micheline Kamber. 2006. *Data Mining: Concepts and Techniques* (2nd ed.). Elsevier, San Francisco, CA.

R. Kohavi. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence* 14, 2, 1137–1145.

K. Kong and M. Tomizuka. 2008. Smooth and continuous human gait phase detection based on foot pressure patterns. *ICRA* 3678–3683.

S. Kullback. 1959. *Information Theory and Statistics*. Wiley, New York, NY.

J. Michael. Fox Foundation for Parkinson's Research. 2016. What causes Parkinson's? Retrieved June 28, 2016 from https://www.michaeljfox.org/understanding-parkinsons/living-with-pd/topic.php?causes.

National Institute of Neurological Disorders and Stroke. 2016. Parkinson's Disease: Hope Through Research. Retrieved June 28, 2016 from http://www.ninds.nih.gov/disorders/parkinsons_disease/detail_parkinsons_disease.htm.

A. Sant'Anna, A. Salarian, and N. Wickstrom. 2011. A new measure of movement symmetry in early Parkinson's disease patients using symbolic processing of inertial sensor data. *IEEE Transactions on Biomedical Engineering* 58, 7, 2127–2135.

C. M. Senanayake and S. A. Senanayake. 2010. Computational intelligent gait-phase detection system to identify pathological gait. *IEEE Transactions on Information Technology in Biomedicine* 14, 5, 1173–1179.

T. Wang, D. Zhang, Z. Wang, J. Jia, H. Ni, and X. Zhou. 2015. Recognizing gait pattern of Parkinson's disease patients based on fine-grained movement function features. *UIC* (2015), to be published.

D. A. Winter, A. E. Patla, J. S. Frank, and S. E. Walt. 1990. Biomechanical walking pattern changes in the fit and healthy elderly. *Physical Therapy* 70, 6, 340–347.

Y. Wu and S. Krishnan. 2010. Statistical analysis of gait rhythm in patients with Parkinson's disease. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 18, 2, 150–158.

F. Yang, D. Espy, and Y. Pai. 2009. Feasible stability region in the frontal plane during human gait. *Annals of Biomedical Engineering* 37, 12, 2604–2614.

G. Yogev, M. Plotnik, C. Peretz, N. Giladi, and J. Hausdorff. 2007. Gait asymmetry in patients with Parkinson's disease and elderly fallers: When does the bilateral coordination of gait require attention? *Experimental Brain Research* 177, 336–346.

G. Yogev, N. Giladi, C. Peretz, S. Springer, E. S. Simon, and J. M. Hausdorff. 2005. Dual tasking, gait rhythmicity, and Parkinson's disease: Which aspects of gait are attention demanding? *European Journal of Neuroscience* 22, 5, 1248–1256.

D. Zhang, Z. Yu, and C. Y. Chin. 2005. Context-aware infrastructure for personalized healthcare. *Studies in Health Technology and Informatics* 117, 154–163.

Received October 2015; revised January 2016; accepted January 2016