Statistical Analysis of Gait Rhythm in Patients With Parkinson's Disease

Yunfeng Wu, *Member, IEEE*, and Sridhar Krishnan, *Senior Member, IEEE*

***Abstract*—To assess the gait variability in patients with Parkinson's disease (PD), we first used the nonparametric Parzen-window method to estimate the probability density functions (PDFs) of stride interval and its two subphases (i.e., swing interval and stance interval). The gait rhythm standard deviation (σ) parameters computed with the PDFs indicated that the gait variability is significantly increased in PD. Signal turns count (STC) was also derived from each outlier-processed gait rhythm time series to serve as a dominant feature, which could be used to characterize the gait variability in PD. Since it was observed that the statistical parameters of swing interval or stance interval were highly correlated with those of stride interval, this article only used the stride interval parameters, i.e., σ_r and STC_r, to form the feature vector in the pattern classification experiments. The results evaluated with the leave-one-out cross-validation method demonstrated that the least squares support vector machine with polynomial kernels was able to provide a classification accurate rate of 90.32% and an area (Az) of 0.952 under the receiver operating characteristic curve, both of which were better than the results obtained with the linear discriminant analysis (accuracy: 67.74%, Az : 0.917). The features and the classifiers used in the present study could be useful for monitoring of the gait in PD.**

***Index Terms*—Gait analysis, movement disorders, Parkinson's disease, Parzen window, probability density function, support vector machine, turns count.**

I. Introduction

SERVING as a pivotal part of the human motor system, the basal ganglia process motor impulses originated from the cerebral cortex and the brain stem, and also send sensory information through the projecting loops in the central nervous system (CNS) [1]. Basal ganglia dysfunction would affect the motor function and may lead to balanced impairment or altered gait rhythm [2]. Parkinson's disease (PD) is a chronic and progressive hypokinetic disorder of the CNS caused by basal ganglia dysfunction. Four major motor symptoms of PD are resting tremor of 4–6 Hz (the most manifest symptom), rigidity (stiffness in muscles), bradykinesia (slow physical movement), and postural instability (loss of postural reflexes) [3]. Other symptoms may include physical fatigue, festination, small shuffling steps, and decreases in both arm swing and walking speed.

To quantify kinetic, spatiotemporal, power spectral, and fractal parameters of the gait in PD, computer-aided analysis has been used in previous studies [4]–[11]. The research group led by Akay, Tamura, and Fujimoto [6], [7] applied the wavelet-based fractal analysis and the time-frequency matching pursuit algorithm to the acceleration signals recorded from PD subjects during climbing stairs and walking along a corridor. Their studies [7] suggested that the acceleration signals of PD patients recorded from one gait cycle to the next would be altered into a more complex pattern, and the fractal dimensions of the body motion tend to be higher in PD. Morris *et al.* [12] observed that PD patients are able to modulate normal walking cadence, but they also reported that the averaged stride length is remarkably shorter in PD patients than in healthy subjects, which has been confirmed in recent studies [13]–[15]. In addition, related studies suggested that the fluctuation dynamics of stride interval are significantly increased in PD patients [5], [10], [16]. Hausdorff *et al.* [10], [15] estimated the stride-to-stride fluctuations in healthy control subjects and in PD patients, respectively. Their results suggested that the coefficient of variation is increased in PD, and also related to the degree of disease severity [10]. They also studied the effects of external cueing using rhythmic auditory stimulation (e.g., by means of a metronome) on gait variability, and the results demonstrated that rhythmic auditory stimulation set to 110% of the step rate may improve mobility and reduce fall risk in PD patients [17]. In addition, Miller *et al.* [18] reported that the electromyographic (EMG) signal variability of gastrocnemius is increased in PD patients.

Although an increase of stride-to-stride variability has been observed in PD, further quantitative studies call for more statistical models that can be used to better characterize the gait variability in PD. In the present study, we used the Parzen-window method to estimate the probability density functions (PDFs) of stride interval and its two subphases (swing interval and stance interval), for healthy subjects and PD patients, respectively. To assess the gait variability, the statistical parameters such as the mean (μ, the expected value computed with the PDF of a specific gait rhythm) and the standard deviation (σ, the square root of the variance computed with the PDF of a specific gait rhythm) were derived from the PDFs obtained. We also measured the signal turns count (STC) parameter to assess the temporal variability of stride interval, swing interval, and stance interval, respectively. It is hypothesized that the σ and STC parameters could be used to characterize the gait variability in PD, and that linear and nonlinear pattern classifiers with these parameters as

Manuscript received April 14, 2009; revised July 03, 2009, September 07, 2009; accepted September 08, 2009. First published October 06, 2009; current version published April 21, 2010. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and in part by the Canada Research Chairs Program. The work of Y. F. Wu was supported by the Faculty Start-Up Research Grant from Xiamen University.

Y. F. Wu is with the Department of Communication Engineering, School of Information Science and Technology, Xiamen University, Xiamen, Fujian 361005, China (e-mail: y.wu@ieee.org).

S. Krishnan is with the Department of Electrical and Computer Engineering, Ryerson University, 350 Victoria Street, Toronto, ON, M5B 2K3 Canada.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNSRE.2009.2033062

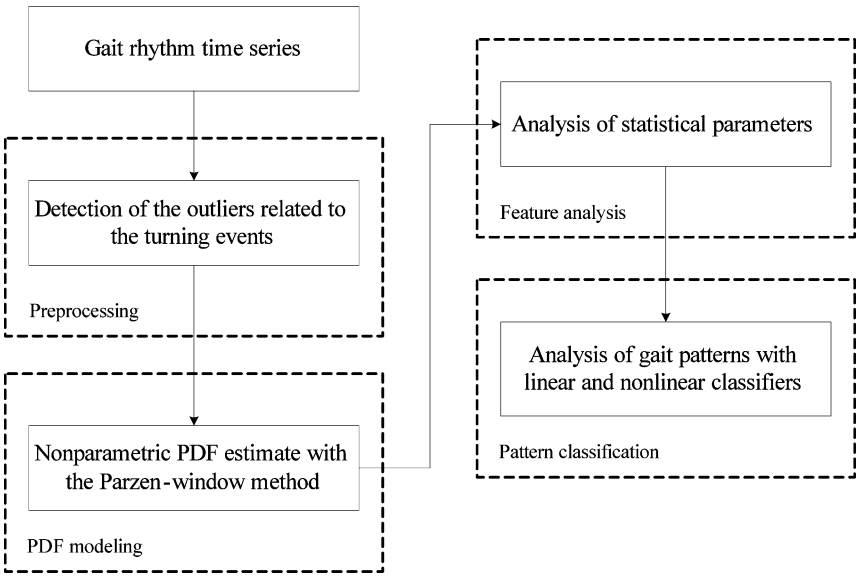

Fig. 1. Block diagram of the procedures for analysis of gait rhythm in PD. PDF: probability density function.

input features could be employed to effectively distinguish the gait patterns of PD patients from those of healthy subjects.

Fig. 1 shows the block diagram of gait analysis procedures in the present study. The procedures include a preprocessing of gait rhythm time series, the nonparametric PDF estimates, the extraction of statistical parameters related to temporal variability of gait rhythm in PD, and the pattern analysis using the linear discriminant analysis (LDA) and the least squares support vector machine (LS-SVM) with nonlinear kernels. The details of these procedures are presented as follows.

II. Data Description

The gait database used in the present study was contributed by Hausdorff *et al.* [10], [19], and is also downloadable via the web page of PhysioNet[1] [20]. Sixteen healthy control (CO) subjects (2 men and 14 women) aged 20–74 years (mean $\pm$ standard deviation, SD[2]: 39.3 ± 18.5 years) were requested to participate. The CO subjects were reported as being free of neurological, respiratory, or cardiovascular disorders. Fifteen PD subjects (10 men and 5 women) aged 44–80 years (mean $\pm$SD : 66.8 ± 10.9 years) were recruited from the Neurology Outpatient Clinic at Massachusetts General Hospital. Age was not independently associated with PD [10]. Either height or weight of the PD subjects was not significantly different from that of the CO subjects.

The Hoehn and Yahr (HAY) score [21] was used to assess the degree of neurological impairment in PD. Two of the PD subjects had a HAY score of 1.5, four had scores of 2 or 2.5, five had a score of 3, and the rest were with severe locomotor impairment ($3 < \mathrm{HAY} < 5$). The PD subjects were established by medical history and examination, and the usage of anti-Parkinsonian medication (a combination of carbidopa and levodopa) for each PD subject was not altered. Table I lists the age of each subject, along with the HAY scores for the PD patients.

TABLE I
Ages of 16 Healthy Control Subjects and 15 Patients With PD. HAY Score Represents the Degree of Neurological Impairment in PD

Healthy control subjects	Age (years)	PD patients	Age (years)	HAY
No. 1	57	No. 1	77	4
No. 2	22	No. 2	44	1.5
No. 3	23	No. 3	80	2
No. 4	52	No. 4	74	3.5
No. 5	47	No. 5	75	2
No. 6	30	No. 6	53	2
No. 7	22	No. 7	64	4
No. 8	22	No. 8	64	4
No. 9	32	No. 9	68	1.5
No. 10	38	No. 10	60	3
No. 11	69	No. 11	74	3
No. 12	74	No. 12	57	3
No. 13	61	No. 13	79	3
No. 14	20	No. 14	57	3
No. 15	20	No. 15	76	2.5
No. 16	40			

All the CO and PD subjects provided informed consent as approved by the Institutional Review Board of the Massachusetts General Hospital. According to the experimental protocol [19], each subject was requested to walk at his or her normal pace along a straight hallway of 77 m in length for 5 min (300 s) without stopping (unless he or she had to turn at the end of the hallway) on level ground. No episode of freezing or dyskinesia was observed in these PD subjects [10]. The force underneath the foot for each stride was recorded using the ultrathin pressure-sensitive switches placed inside each subject's shoes. A recorder (dimensions: $5.5 \times 2 \times 9$ cm; weight: 0.1 kg) was worn on the ankle cuff of each foot and held in place with a wallet on the ankle [22]. The signal recorded was digitized by an on-board analog-to-digital converter at the sampling rate of 300 Hz with

[1]Online available at http://www.physionet.org/physiobank/database/gaitndd/

[2]For a clear presentation, the terms of standard deviation used to characterize the gait variability was mentioned as σ, and the abbreviation of standard deviation, i.e., SD, was used to describe the statistical dispersion of a random variable.

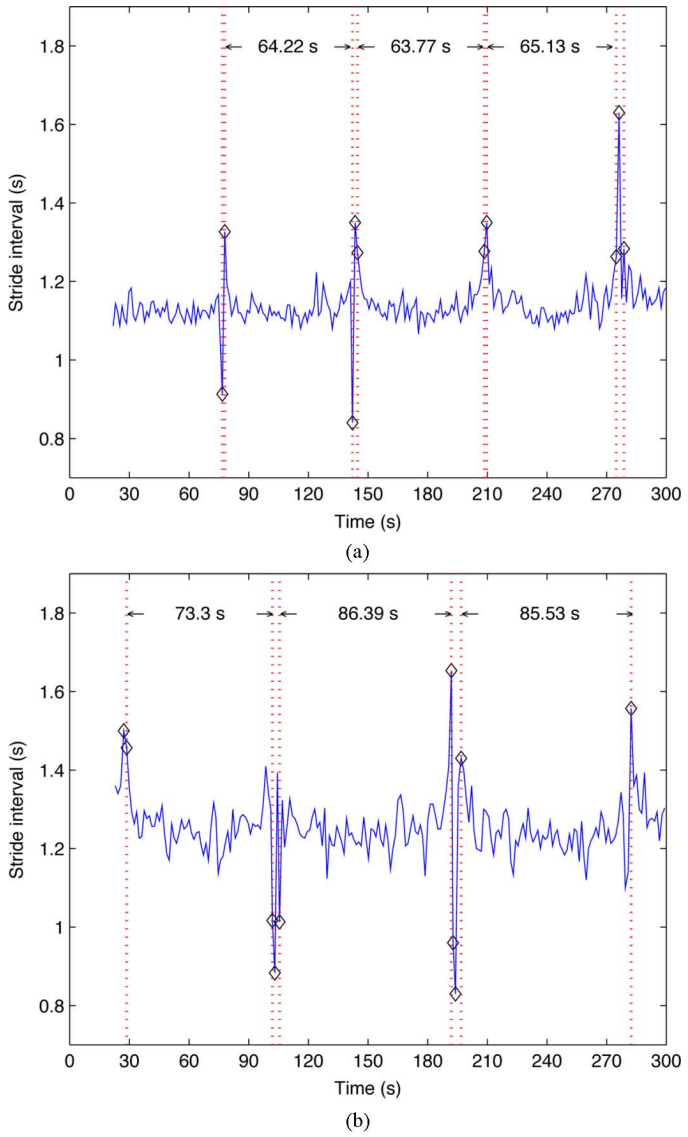

Fig. 2. Examples of the raw stride interval time series: (a) of a 74-year-old male control subject; (b) of an age-matched female subject with PD (severe impairment). The outliers marked with *diamonds* represent the data samples that were two times standard deviations greater or less than the median value of stride interval over the entire time series.

12-bit resolution per sample. In the present study, we only focused on the right-foot gait rhythm in terms of stride interval (time from initial contact of one foot to the subsequent contact of the same foot), swing interval (amount of time one foot is in the air), and stance interval (time of bilateral foot contact).

III. Detection of Outliers

To minimize the startup effects in the gait acquisition, the samples of each gait rhythm time series recorded in the first 20 s were removed before the statistical analysis (see Fig. 2), which was the same as implemented in previous studies of Hausdorff *et al.* [10], [19]. During the 5-min walking period, every time the subjects reached the end of the hallway, they had to turn around and then continue walking [19]. Therefore, the strides recorded during these walking turns should be regarded as outliers. According to the "three-sigma rule" [23], about 95% and 99.7% of the normally distributed probability values lie within 2-SD and 3-SD distances from the mean, respectively. In the present study, we detected the outliers that were 2 SDs greater or less than the median value over the entire stride interval time series. We did not choose the value of 3 SDs, because it failed to detect some visible outliers due to the walking turns. Fig. 2 illustrates the detection of outliers present in the stride interval time series of a 74-year-old PD patient (PD No. 4) and of an age-matched CO subject (control No. 12), respectively. According to the outlier annotations, a few 77-m straight walk periods may be segmented, as shown in Fig. 2. Thus, the averaged walking speed of the CO and PD subjects illustrated in Fig. 2 can be derived as $(3 \times 77)/(64.22 + 63.77 + 65.13) = 1.2$ m/s and $(3 \times 77)/(73.3 + 86.39 + 85.53) = 0.94$ m/s, respectively. These values of gait speed were close to those in the corresponding clinical records (CO: 1.26 m/s; PD: 0.91 m/s), which demonstrated that such an outlier detection procedure was reliable.

For further gait analysis, the stride interval outliers were replaced with the median value of the stride interval time series. We considered the median value instead of the mean of a gait rhythm time series because we observed that some outliers possessed very large values, and might affect the mean of the entire time series (see Fig. 1). Similarly, with the annotations of the stride interval outliers, the samples of swing interval and stance interval related to the walking turns at the end of the hallway were also replaced with the median values of swing interval and stance interval in the corresponding time series, respectively.

IV. Statistical Analysis of Gait Variability

A. Nonparametric Estimate of Gait Rhythm PDFs

To obtain the PDF models, we first computed the histogram for each gait rhythm x, which was referred to as the stride interval or its subphases. The histogram of the gait rhythm was established with B bins, which helped calculate the probability of occurrence with B containers of equal length in the amplitude range of x. The optimal number of bins with regard to a specific gait rhythm was determined according to the minimization criterion of the global mean squared error between the Gaussian density function and the histogram estimated, which is also referred to as Scott's choice [24], i.e.,

$$B = \frac{(x^h - x^l)}{3.49sn^{-1/3}} \tag{1}$$

where s and n denote the SD and the number of samples in a specific gait rhythm time series, respectively; x^h and x^l represent the highest and the lowest values in the amplitude range of x, respectively.

The Parzen-window approach was applied to obtain a nonparametric estimate of the PDF of a specific gait rhythm from the samples in the corresponding time series. Consider the situation where a gait rhythm time series contains a set of samples sorted by amplitude, $Z = \{z_1, z_2, \ldots, z_K\}$, with an unknown PDF $p(z)$. The nonparametric estimate of the underlying PDF from Z is provided by the function

$$\hat{p}(z) = \frac{1}{K}\sum_{k=1}^{K} w(z - z_k) \tag{2}$$

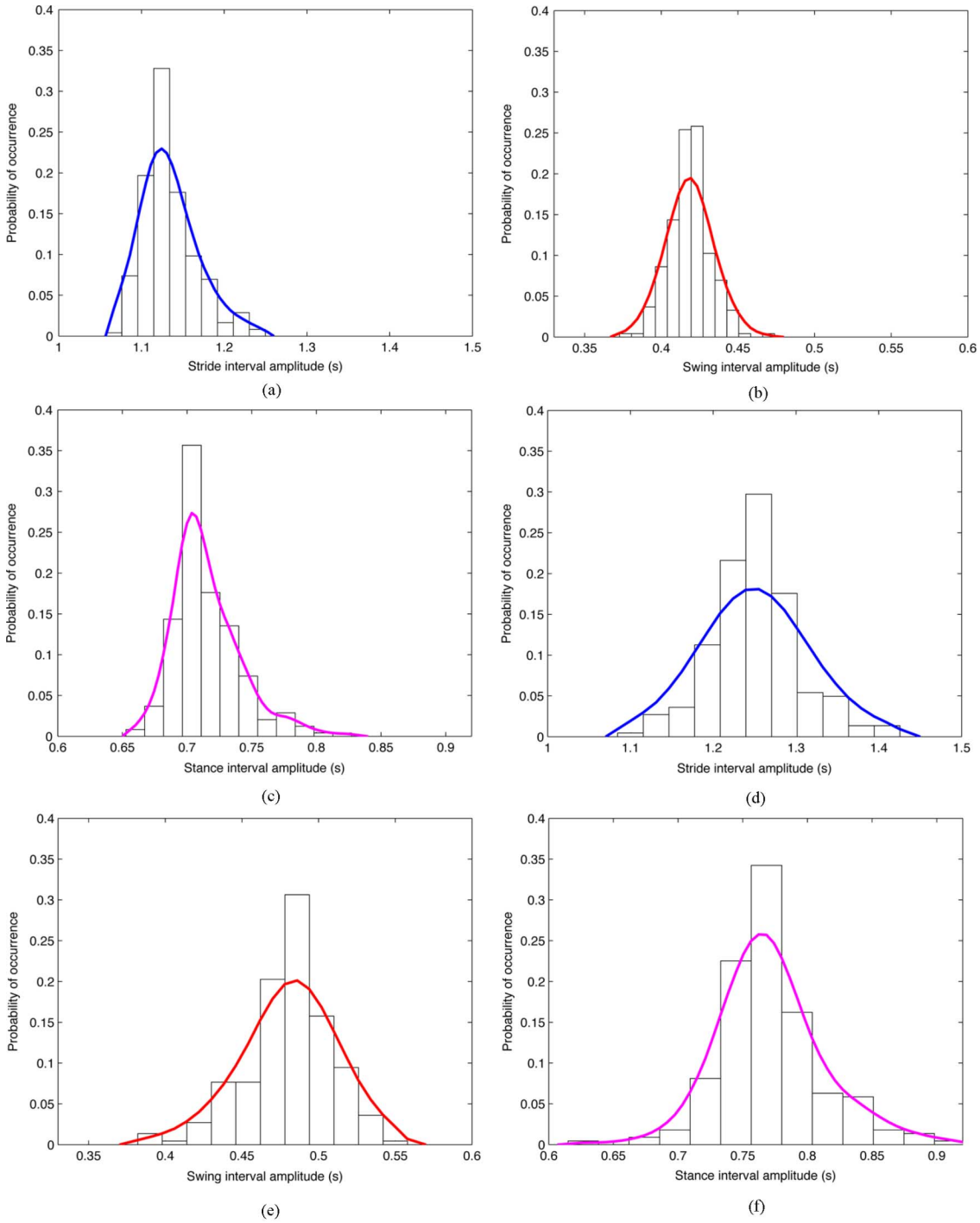

Fig. 3. Nonparametric Parzen-window estimates of the gait rhythm PDF models of the CO and PD subjects in Fig. 2: the PDFs of (a) stride interval ($\mu_r = 1.13$ s, $\sigma_r = 0.03$ s, $B = 10$ bins, $\sigma_{\rm P} = 0.02$), (b) swing interval ($\mu_w = 0.42$ s, $\sigma_w = 0.01$ s, $B = 13$ bins, $\sigma_{\rm P} = 0.01$), and (c) stance interval ($\mu_a = 0.72$ s, $\sigma_a = 0.03$ s, $B = 12$ bins, $\sigma_{\rm P} = 0.01$) in the CO subject (aged 74); and the PDFs of (d) stride interval ($\mu_r = 1.25$ s, $\sigma_r = 0.05$ s, $B = 11$ bins, $\sigma_{\rm P} = 0.05$), (e) swing interval ($\mu_w = 0.48$ s, $\sigma_w = 0.03$ s, $B = 11$ bins, $\sigma_{\rm P} = 0.02$), and (f) stance interval ($\mu_a = 0.77$ s, $\sigma_a = 0.04$ s, $B = 13$ bins, $\sigma_{\rm P} = 0.02$) in the PD subject (aged 74, severe impairment).

where $w(\cdot)$ is a window function that integrates to unity. In the present study, we considered the Gaussian window, which may be expressed as

$$w(z - z_k) = \frac{1}{\sigma_{\rm P}\sqrt{2\pi}} \exp\left[-\frac{(z - z_k)^2}{2\sigma_{\rm P}^2}\right] \qquad (3)$$

where $\sigma_{\rm P}$ is the spread parameter that determines the width of a Gaussian window, the center of which is located at z_k. The value of the spread parameter was varied over the range $[0.01, 0.1]$, with an incremental step of 0.01, to carry out each experiment for a specific gait rhythm. Following the criterion of minimization of the mean-squared error between the estimated PDF and the histogram, the optimal values of $\sigma_{\rm P}$ to fit the PDFs of stride interval were 0.02 and 0.05 for the CO and PD subjects, respectively; the optimal $\sigma_{\rm P}$ to fit the PDFs of swing interval were 0.01 and 0.02 for the CO and PD subjects, respectively; the optimal

σ_{P} to fit the PDFs of stance interval were 0.01 and 0.02 for the CO and PD subjects, respectively. Fig. 3 shows the histogram graphics along with the estimated PDF curves for the stride interval and its two subphases in Fig. 2. It can be observed that the PDFs of the 74-year-old PD patient shown in Fig. 3(d)–(f) were more dispersed than those of the age-matched CO subject shown in Fig. 3(a)–(c).

Based on the estimated gait rhythm PDF, denoted by $\hat{p}(x_b)$, with $x_b, b = 1, 2, \ldots, B$, indicating the B bins used to represent the amplitude range of x, the mean μ and variance σ^2 can be derived as

$$\mu = \sum_{b=1}^{B} x_b \hat{p}(x_b) \tag{4}$$

$$\sigma^2 = \sum_{b=1}^{B} (x_b - \mu)^2 \hat{p}(x_b). \tag{5}$$

In this paper, the parameters of μ and σ are presented as μ_r and σ_r for stride interval, μ_w and σ_w for swing interval, and μ_a and σ_a for stance interval, respectively. The statistics of μ and σ computed for stride interval and its two subphases of the CO subjects and PD patients are presented in Table II, and also illustrated in Fig. 4.

The Student's t-test [25] was utilized in our experiments to test whether or not the mean values of the statistical parameters of the PD patients significantly differ from those associated with the CO subjects. According to the p-values listed in Table II, any of the stride interval and its two subphases in the PD subjects possessed the value of μ similar to that of the corresponding gait rhythm in the CO subjects, as shown in Fig. 4(a). However, the p-values of the σ parameters demonstrated that the variations of gait rhythm in PD patients became significantly higher than those in CO subjects, which supported the observation of Hausdorff *et al.* [10]. In fact, from Fig. 4(b), we may observe that the σ bars of PD patients were about two times as high as those of CO subjects.

B. Signal Turns Count

In a given time series, $\{x(i)\}, i = 1, 2, \ldots, I$, a data sample can be identified as a signal "turn" if it satisfies the following two conditions at the same time [26]: 1) it represents an alteration in the direction, i.e., a change in the sign of the derivative (from positive to negative, or vice versa); 2) the difference (absolute value) between the amplitude of the current sample and that of the preceding sample should be greater than a specific threshold. The detection of a signal turn may be expressed as

$$x(i) \text{ is a signal turn} \quad \text{if} \begin{cases} [x(i) - x(i-1)][x(i+1) - x(i)] < 0 \\ |x(i+1) - x(i)| \geq \mathrm{Th}, \ 2 \leq i \leq I - 1 \end{cases} \tag{6}$$

where Th denotes the threshold. The number of signal turns in a time series represents the degree of signal variability.

Willison [27] first used the signal turns count (STC) method to analyze the EMG signal, and the experiments showed that the EMG signal recorded from a patient with a myopathy usually possesses more signal turns than that of a normal subject. The study of Rangayyan and Wu [28] on the knee-joint vibroarthrographic signal analysis suggested that the value of STC could

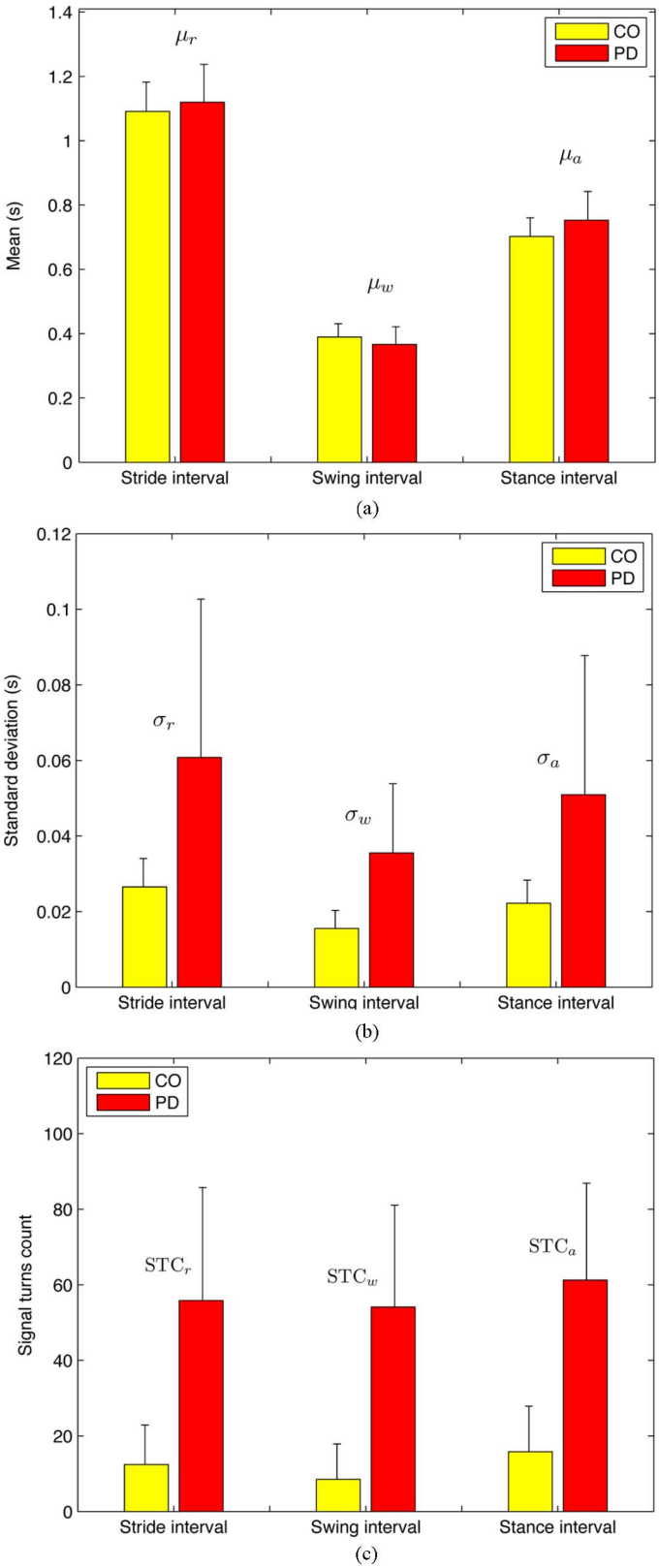

Fig. 4. Bar graphics of the mean value of (a) μ, (b) σ, and (c) STC parameters with regard to the stride interval, swing interval, and stance interval in the CO subjects and the PD patients. Vertical lines indicate SD. The value of each bar is listed in Table II.

TABLE II
MEAN, SD, AND SIGNIFICANCE OF SEPARABILITY (STUDENT'S t-TEST p-VALUE) OF EACH STATISTICAL PARAMETER COMPUTED FOR 16 HEALTHY CONTROL (CO) SUBJECTS AND 15 PATIENTS WITH PD

Gait rhythm	Statistical parameters	CO subjects Mean ± SD	PD patients Mean ± SD	p-values† CO vs. PD
Stride interval	μ_r (s)	1.09 ± 0.09	1.12 ± 0.12	0.45
	σ_r (s)	0.03 ± 0.01	0.06 ± 0.04	< 0.01
	STC_r	12.44 ± 10.46	55.87 ± 29.89	< 0.01
Swing interval	μ_w (s)	0.39 ± 0.04	0.37 ± 0.05	0.21
	σ_w (s)	0.02 ± 0.01	0.04 ± 0.02	< 0.01
	STC_w	8.5 ± 9.39	54.13 ± 26.96	< 0.01
Stance interval	μ_a (s)	0.7 ± 0.06	0.75 ± 0.09	0.07
	σ_a (s)	0.02 ± 0.01	0.05 ± 0.04	< 0.01
	STC_a	15.81 ± 12.07	61.27 ± 25.62	< 0.01

†: $p < 0.01$, significant.

be used as a dominant feature for screening of knee-joint disorders. Wu and Krishnan [29] recently applied the STC method to measure the gait fluctuations in amyotrophic lateral sclerosis. Their results indicated that the degree of swing interval fluctuations in patients with amyotrophic lateral sclerosis is significantly higher than that in healthy controls [29]. Although the aforementioned studies suggested that the STC method is very suited for the variability analysis of physiological signals, there lacks an application of pattern analysis of the gait in PD. In the present study, we would like to use the STC method to characterize the variability of the gait in PD patients.

In the experiments, the STC values were computed for the outlier-processed gait rhythm time series of the CO and PD subjects, respectively. Since the outliers of stride interval, along with its two subphases, were replaced with the median value over the corresponding time series, there would be no change in amplitude in these modified samples, so that the STC values would not be affected by the procedure of the outlier processing described in Section III. The STC threshold was varied over the range $[0.01, 5]$, with an incremental step of 0.01 s. The optimal STC thresholds (i.e., 0.07, 0.05, and 3.54 s for the time series of stride interval, swing interval, and stance interval, respectively) were determined according to the most significance of separability characterized by the p-values. Fig. 5 illustrates the signal turns (*asterisks*) detected in the outlier-preprocessed stride interval time series in Fig. 2. It is clear that the stride interval time series of the PD subject possessed a larger number of signal turns that were greater than the threshold of 0.07 s. According to Table II, the STC parameters of the gait rhythm in PD patients were considerably different from those in CO subjects. In Fig. 4(c), we may observe that the STC bars of the gait rhythm in PD patients were four times as high as those in CO subjects.

C. Correlation Analysis of Statistical Parameters

The t-test results showed that both of the σ and STC parameters were significantly increased in PD, so that these gait variability parameters could be considered as the feature candidates for further pattern classifications. With the aim to select the representative features, we computed the correlations of the two types of statistical parameters using Pearson's correlation coefficient.

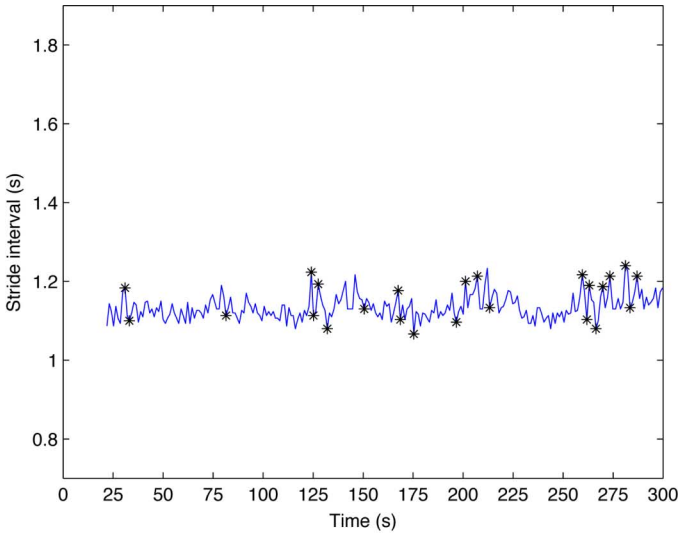

Fig. 5. Detection of the STC in the outlier-processed stride interval time series: (a) of a 74-year-old male control subject; (b) of an age-matched female patient with PD (severe impairment). Significant turns detected have been marked with *asterisks*. The 2-SD outliers had been replaced with the median value of stride interval over the entire time series.

TABLE III
CORRELATION COEFFICIENTS BETWEEN THE σ PARAMETERS OF STRIDE INTERVAL, SWING INTERVAL, AND STANCE INTERVAL

σ parameter	σ_r	σ_w	σ_a
σ_r	1	0.99	0.94
σ_w	0.99	1	0.94
σ_a	0.94	0.94	1

The correlation matrices of the σ and STC parameters are tabulated in Tables III and IV, respectively. It is worth noting that both of these two types of gait variability parameters exhibited high correlations between any two of the stride interval and its two subphases. Based on such observations, we only selected the σ_r and STC_r parameters to form the feature vector $(\sigma_r, \mathrm{STC}_r)^T$ for the pattern analysis, as the other parameters

TABLE IV
CORRELATION COEFFICIENTS BETWEEN THE STC PARAMETERS OF STRIDE INTERVAL, SWING INTERVAL, AND STANCE INTERVAL

STC parameter	STC_r	STC_w	STC_a
STC_r	1	0.94	0.81
STC_w	0.94	1	0.94
STC_a	0.81	0.94	1

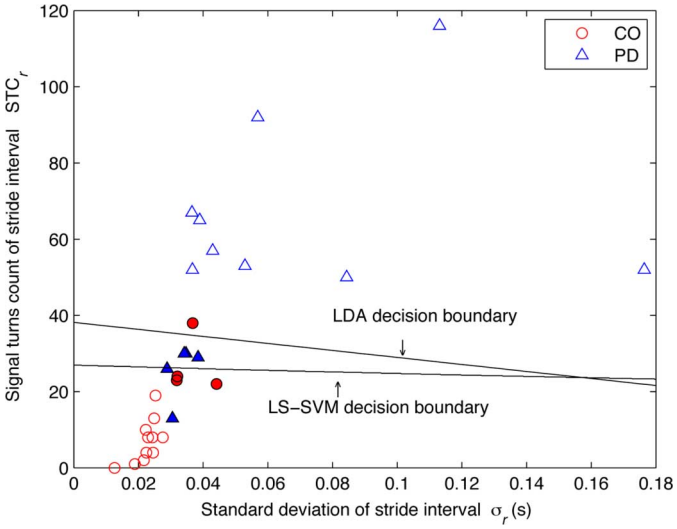

Fig. 6. Scatter plot of the gait patterns of the CO subjects, marked as *circles*, and of the patients with PD, marked as *triangles*, in the 2-D feature space. The decision boundaries were provided by the LDA and the LS-SVM, respectively. Both of these two classifiers were trained with the input features of all 16 healthy control subjects and 15 PD patients, rather than the leave-one-out cross-validation method. Solid data points indicate the support vectors.

had the linear relationships with these two parameters. Fig. 6 plots in the $(\sigma_r, \mathrm{STC}_r)^T$ feature space the gait patterns associated with the CO subjects and the PD patients, respectively. It can be observed that the CO patterns mainly congregated in the area where $\sigma_r < 0.04$ s and $\mathrm{STC}_r < 40$, whereas most of the PD patterns were widely dispersed in the area where $\sigma_r > 0.03$ s and $\mathrm{STC}_r > 20$.

V. PATTERN CLASSIFICATION METHODS AND RESULTS

In the present study, we implemented the linear discriminant analysis (LDA) [30] and the least squares support vector machine (LS-SVM) proposed by Suykens *et al.* [31], for the linear and nonlinear pattern classifications, respectively. The LS-SVM is a reformulation to the standard support vector machine (SVM) proposed by Cortes and Vapnik [32]. Similar to the standard SVM, the LS-SVM is also a type of kernel-based feedforward network, the learning of which follows the principle of structural risk minimization [32]. To optimize the model parameters of the LS-SVM, a subset of the representative training data is selected to serve as the support vectors, which are considered to be the most informative for the classification task. By choosing the nonlinear inner-product kernels in the network, the LS-SVM is able to perform the same function as the polynomial learning machine (with polynomial kernels), radial basis function network (with Gaussian kernels), or multilayer perceptron with a single hidden layer (with sigmoid

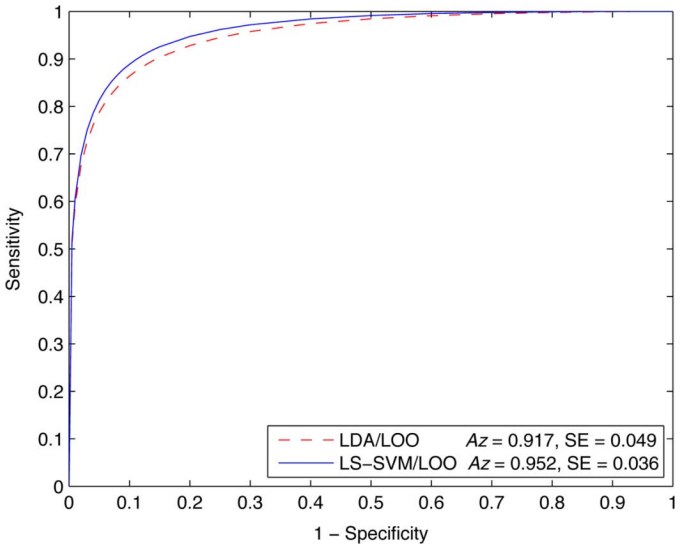

Fig. 7. ROC curves obtained with the LDA and the LS-SVM with polynomial kernels. The features to the inputs of both the LDA and the LS-SVM were σ_r and STC_r, and the classification performance of the two classifiers was evaluated with the LOO cross-validation method. SE: standard error of the area (Az) under the ROC curve.

kernels) [30]. The major advantage of the LS-SVM is that the learning process is implemented by minimizing a regularized least squares cost function with equality constraints under the Karush–Kuhn–Tucker (KKT) condition, which enables the LS-SVM to perform more efficiently than the standard SVM. For more details of the LS-SVM theoretical framework, readers are referred to the book of Suykens *et al.* [31].

In addition to the classification accuracy presented in percentage, receiver operating characteristic (ROC) curve was derived to serve as an additional measure of the overall diagnostic performance in each classification experiment. The classification performance of our experiments were evaluated with the leave-one-out (LOO) cross-validation method [30], and the ROC curves were generated using the software ROCKIT[3] provided by the University of Chicago [33]. We compared the classification accuracy and the area (Az) under the ROC curve obtained by the LS-SVM with different types of nonlinear kernels and different values of the kernel parameters, and chose the polynomial kernels, which helped the LS-SVM provide the best diagnostic performance. The polynomial kernel function $\kappa(\cdot)$ used in the experiments can be expressed as

$$\kappa(\mathbf{f}_i, \mathbf{f}_j) = \left(\mathbf{f}_i^T \mathbf{f}_j + m\right)^d \tag{7}$$

where $\mathbf{f}_i$ represents the feature vector $(\sigma_r, \mathrm{STC}_r)^T$ associated with the ith subject, $d = 2$ is the degree of the polynomial kernel, and $m = 1$ denotes the intercept.

In the classification experiments, the LS-SVM evaluated with the LOO method, labeled as LS-SVM/LOO, provided an overall accurate rate of 90.32%, which was remarkably better than the LDA/LOO (accuracy: 67.74%). As shown in Fig. 7, the ROC curves showed that the LS-SVM/LOO was able to provide the ROC curve (Az : 0.952; standard error, SE: 0.036) consistently over that obtained with the LDA/LOO (Az : 0.917; SE: 0.049).

[3]Online available at http://www-radiology.uchicago.edu/krl/KRL_ROC/software_index6.htm

In each test of the LOO cross-validation method, only one gait pattern was chosen for the validation test, and meanwhile, the remaining patterns were used for the training purpose. Such that the support vectors and the decision boundary obtained with the LS-SVM may be altered from one experiment to another. Fig. 6 plots the support vectors (solid *circles* and *triangles*) generated by the LS-SVM, along with the decision boundaries provided by the linear and nonlinear classifiers trained with the patterns of all 31 subjects. We may observe that the LDA trained with the whole data set reached an accurate rate of 80.65%, much higher than the 67.74% accuracy evaluated with the LOO cross-validation method. It is clear that the linear classifier did not have a good generalization ability when solving the gait pattern classification task. On the other hand, the LS-SVM, based on the proper support vectors, was able to generate a better discriminant function to identify the PD patterns. It is worth noting that the support vectors were highly concentrated within the small ranges of σ_r and STC_r, so that the entire decision boundary of the LS-SVM looked quite similar to an oblique line in Fig. 6. In addition, the STC_r parameter provided more discriminant information than the σ_r parameter, because the decision boundaries of the LDA and LS-SVM appeared much closer to the horizontal line, rather than the vertical line.

VI. Discussion and Conclusion

The present study demonstrated that the gait variability, in terms of statistical parameters of stride interval (i.e., σ_r and STC_r), would be increased in PD. But it is worth noting that the present study has some limitations because the subject groups of healthy controls and PD patients were not matched with respect to age. Although the average age of the PD patients (66.8 years old) was much older than that of the healthy CO subjects (39.3 years old), the study of Hausdorff *et al.* [34] suggested that the effects of neurological diseases are more pronounced than the those due to aging. In the present study, there was also some evidence indicating that the gait variability would not only be altered due to physiological aging, but also as a result of pathology. For example, a 57-year-old PD patient (PD No. 14) possessed the STC_r value of 114 (the largest among those of all 31 subjects), but two PD patients at the age of 68 and 76 (PD No. 9 and 15, which were included in the set of support vectors in Fig. 6) possessed the STC_r values of 29 and 26, respectively. Nevertheless, it would be better to clarify the effects of aging on the gait in the future work.

Since the STC parameter represents the sum of the signal turns in a gait rhythm time series, the final count is dependent upon the number of strides of each subject. The results showed that the averaged stride interval (in terms of μ_r) in PD patients was similar to that in control subjects, which implied that, within the same monitoring period (5 min in the present study), the number of strides of a PD patient was not significantly different from that of a control subject. Hence, the STC method for the analysis of the gait in PD patients became reasonable.

In the experiments, the p-values of σ_r and STC_r indicated that these two statistical parameters may be extracted as the dominant features for the analysis of PD gait patterns. The nonlinear classifier based on the σ_r and STC_r parameters can correctly identify over 90% of the 31 subjects studied, which is comparable to the results obtained using the multiple logistic regression analysis (with double support time coefficient of variation and walking speed as input features) [10]. It is positively considered that these two features could be potentially used in rehabilitation related applications such as gait monitoring.

The same gait database was also investigated in several recent studies. Carletti *et al.* [35] proposed a linear model to interpret the stride interval time series in PD. However, such a linear model is only suited for walking at a constant speed. About 70%, in particular 213 s (PD) and 220 s (CO), of the stride interval time series can be correctly predicted with this model. Dutta *et al.* [36] extracted three types of cross-correlation-based features (i.e., centroid, mean-square abscissa, and variance of the abscissa) from the double support time, stance interval, and swing interval time series, and then applied the Elman's recurrent neural network to distinguish the gait patterns in neurodegenerative diseases (PD, Huntington's disease, and amyotrophic lateral sclerosis) from those of healthy subjects. Although the gait rhythm features illustrated in their paper showed notable correlations between each other, they did not implement the feature-correlation analysis, which is considered to be necessary before the classifications, because we have observed that the statistical parameters of swing interval or stance interval were highly correlated with those of stride interval. Zheng *et al.* [37] simply sent ten raw gait rhythm time series to the classifiers, which were evaluated with the 10-fold cross-validation. Their results showed that the standard SVM (accuracy: 86.43%; Az : 0.92, PD versus CO) outperformed the other classifiers. Our classification experiments demonstrated that the LS-SVM could also be a good choice for the analysis of PD gait patterns, as the LS-SVM classifier with polynomial kernels can provide a decision boundary better than the LDA. Compared with the study of Zheng *et al.* [37], we found it better to implement the feature extraction in the gait pattern classifications, because it was clear that the statistical parameters and the STC features contained informative information and can help improve the classification performance.

It would also be interesting to study in the further work the effects of treadmill walking [38] or external cueing [17] on the stride-to-stride fluctuations, by using the gait rhythm PDF modeling and the STC method.

Acknowledgment

The authors would like to express gratitude toward the anonymous reviewers whose valuable comments and suggestions greatly improved the quality of the paper.

References

[1] J. Sian, M. Gerlach, M. B. H. Youdim, and P. Riederer, "Parkinson's disease: A major hypokinetic basal ganglia disorder," *J. Neural Transmission*, vol. 106, no. 5–6, pp. 443–476, 1999.

[2] C. D. Marsden, "The mysterious motor function of the basal ganglia: The Robert Wartenberg lecture," *Neurology*, vol. 32, no. 5, pp. 514–539, 1982.

[3] J. Jankovic, "Parkinson's disease: Clinical features and diagnosis," *J. Neurol., Neurosurgery, Psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[4] M. E. Morris, J. McGinley, F. Huxham, J. Collier, and R. Iansek, "Constraints on the kinetic and spatiotemporal parameters of gait in Parkinson's disease," *Human Movement Sci.*, vol. 18, no. 2–3, pp. 461–483, 1999.

[5] O. Blin, A. M. Ferrandez, and G. Serratrice, "Quantitative analysis of gait in Parkinson patients: Increased variability of stride length," *J. Neurological Sci.*, vol. 98, no. 1, pp. 91–97, 1990.
[6] M. Sekine, T. Tamura, M. Akay, T. Fujimoto, T. Togawa, and Y. Fukui, "Discrimination of walking patterns using wavelet-based fractal analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 10, no. 3, pp. 188–196, Sep. 2002.
[7] M. Sekine, M. Akay, T. Tamura, and Y. Higashi, "Fractal dynamics of body motion in patients with Parkinson's disease," *J. Neural Eng.*, vol. 1, no. 1, pp. 8–15, 2004.
[8] J. M. Hausdorff, C. K. Peng, Z. Ladin, J. Y. Wei, and A. L. Goldberger, "Is walking a random walk? Evidence for long-range correlations in stride interval of human gait," *J. Appl. Physiol.*, vol. 78, no. 1, pp. 349–358, 1995.
[9] J. M. Hausdorff, P. L. Purdon, C. K. Peng, Z. Ladin, J. Y. Wei, and A. L. Goldberger, "Fractal dynamics of human gait: Stability of long-range correlations in stride interval fluctuations," *J. Appl. Physiol.*, vol. 80, no. 5, pp. 1448–1457, 1996.
[10] J. M. Hausdorff, M. E. Cudkowicz, R. Firtion, J. Y. Wei, and A. L. Goldberger, "Gait variability and basal ganglia disorders: Stride-to-stride variations of gait cycle timing in Parkinson's disease and Huntington's disease," *Movement Disorders*, vol. 13, no. 3, pp. 428–437, 1998.
[11] J. M. Hausdorff and N. B. Alexander, *Gait Disorders: Evaluation and Management*. New York: Informa Healthcare, 2005.
[12] M. E. Morris, R. Iansek, T. A. Matyas, and J. J. Summers, "Ability to modulate walking cadence remains intact in Parkinson's disease," *J. Neurol., Neurosurgery Psychiatry*, vol. 57, no. 12, pp. 1532–1534, 1994.
[13] M. E. Morris, T. A. Matyas, R. Iansek, and J. J. Summers, "Temporal stability of gait in Parkinson's disease," *Physical Therapy*, vol. 76, no. 7, pp. 763–777, 1996.
[14] M. E. Morris, F. Huxham, J. McGinley, K. Dodd, and R. Iansek, "The biomechanics and motor control of gait in Parkinson disease," *Clin. Biomechan.*, vol. 16, no. 6, pp. 459–470, 2001.
[15] J. M. Hausdorff, "Gait dynamics, fractals and falls: Finding meaning in the stride-to-stride fluctuations of human walking," *Human Movement Sci.*, vol. 26, no. 4, pp. 555–589, 2007.
[16] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, and J. M. Hausdorff, "Effect of gait speed on gait rhythmicity in Parkinson's disease: Variability of stride time and swing time respond differently," *J. Neuroeng. Rehabilitat.*, vol. 2, p. 23, 2005.
[17] J. M. Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger, C. Peretz, and N. Giladi, "Rhythmic auditory stimulation modulates gait variability in Parkinson's disease," *Eur. J. Neurosci.*, vol. 26, no. 8, pp. 2369–2375, 2007.
[18] R. A. Miller, M. H. Thaut, G. C. McIntosh, and R. R. Rice, "Components of EMG symmetry and variability in parkinsonian and healthy elderly gait," *Electroencephalogr. Clin. Neurophysiol.*, vol. 101, no. 1, pp. 1–7, 1996.
[19] J. M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, A. L. Peterson, D. Kaliton, and A. L. Goldberger, "Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis," *J. Appl. Physiol.*, vol. 88, no. 6, pp. 2045–2053, 2000.
[20] G. B. Moody, R. G. Mark, and A. L. Goldberger, "PhysioNet: A web-based resource for the study of physiologic signals," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 70–75, May/Jun. 2001.
[21] M. M. Hoehn and M. D. Yahr, "Parkinsonism: Onset, progression, and mortality," *Neurology*, vol. 17, no. 5, pp. 427–442, 1967.
[22] J. M. Hausdorff, Z. Ladin, and J. Y. Wei, "Footswitch system for measurement of the temporal parameters of gait," *J. Biomechan.*, vol. 28, no. 3, pp. 347–351, 1995.
[23] G. J. Hahn and S. S. Shapiro, *Statistical Models in Engineering*. New York: Wiley, 1994.
[24] D. W. Scott, "On optimal and data-based histograms," *Biometrika*, vol. 66, no. 3, pp. 605–610, 1979.
[25] J. F. Box, "Gosset, fisher, and the t distribution," *Am. Stat.*, vol. 35, no. 2, pp. 61–66, 1981.
[26] R. M. Rangayyan, *Biomedical Signal Analysis: A Case-Study Approach*. New York: IEEE Wiley, 2002.
[27] R. G. Willison, "Analysis of electrical activity in healthy and dystrophic muscle in man," *J. Neurol., Neurosurgery, Psychiatry*, vol. 27, no. 5, pp. 386–394, 1964.
[28] R. M. Rangayyan and Y. F. Wu, "Analysis of vibroarthrographic signals with features related to signal variability and radial-basis functions," *Ann. Biomed. Eng.*, vol. 37, no. 1, pp. 156–163, 2009.
[29] Y. F. Wu and S. Krishnan, "Computer-aided analysis of gait rhythm fluctuations in amyotrophic lateral sclerosis," *Med. Biol. Eng. Comput.*, vol. 47, no. 11, pp. 1165–1171, 2009.
[30] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
[31] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*. Singapore: World Scientific Publishing, 2002.
[32] C. Cortes and V. N. Vapnik, "Support-vector networks," *Machine Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[33] C. Metz, B. Herman, and J. Shen, "Maximum-likelihood estimation of ROC curves from continuously-distributed data," *Stat. Med.*, vol. 17, no. 9, pp. 1033–1053, 1998.
[34] J. M. Hausdorff, S. L. Mitchell, R. Firtion, C. K. Peng, M. E. Cudkowicz, J. Y. Wei, and A. L. Goldberger, "Altered fractal dynamics of gait: Reduced stride-interval correlations with aging and Huntington's disease," *J. Appl. Physiol.*, vol. 82, no. 1, pp. 262–269, 1997.
[35] T. Carletti, D. Fanelli, and A. Guarino, "A new route to non invasive diagnosis in neurodegenerative diseases?," *Neurosci. Lett.*, vol. 394, no. 3, pp. 252–255, 2006.
[36] S. Dutta, A. Chatterjee, and S. Munshi, "An automated hierarchical gait pattern identification tool employing cross-correlation-based feature extraction and recurrent neural network based classification," *Expert Syst.*, vol. 26, no. 2, pp. 202–217, 2009.
[37] H. R. Zheng, M. J. Yang, and S. McClean, "Machine learning and statistical approaches to support the discrimination of neuro-degenerative diseases based on gait analysis," in *Intelligent Patient Management*. Berlin, Germany: Springer, 2009, pp. 57–70.
[38] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, and J. M. Hausdorff, "Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson's disease," *Movement Disorders*, vol. 20, no. 9, pp. 1109–1114, 2005.

Yunfeng Wu (S'01–M'09) received the B.E. and Ph.D. degrees from Beijing University of Posts and Telecommunications (BUPT), China, in 2003 and 2008, respectively. He was a Croucher visiting scholar at The Open University of Hong Kong (OUHK), Hong Kong, and then worked as a Post-Doctoral Fellow at the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, Canada.

He is currently an Associate Professor at the Department of Communication Engineering, School of Information Science and Technology, Xiamen University, Xiamen, Fujian Province, China. His research interests include biomedical signal and image processing, pattern recognition, and neural signal processing.

Dr. Wu received the travel grants of ICONIP'06, ISSS-MDBS'06, and IJCNN'04, and was the recipient of Second Place of the 2002 IEEE IAS Myron Zucker Student Design Contest; Meritorious award of the 2002 Interdisciplinary Contest in Modeling; and the IBM Chinese Excellent Students Scholarship in 2002. Dr. Wu is a member of BMES and INNS.

Sridhar Krishnan (S'94–M'99–SM'05) received the B.E. degree from Anna University, Chennai, India, in 1993, and the M.Sc. and Ph.D. degrees from the University of Calgary, Calgary, AB, Canada, in 1996 and 1999, respectively.

He joined Ryerson University, Toronto, ON, Canada, in 1999, and is currently a Professor and Canada Research Chair in Biomedical Signal Analysis. His research interests include adaptive signal representations and their applications in biomedicine and multimedia.

Prof. Krishnan is the founding Chair of the IEEE Signal Processing Chapter of Toronto Section, and is a registered Professional Engineer of the Province of Ontario.