Journal of Biomechanics 47 (2014) 186–192

Contents lists available at ScienceDirect

Journal of Biomechanics

journal homepage: www.elsevier.com/locate/jbiomech
www.JBiomech.com

Statistical method for prediction of gait kinematics with Gaussian process regression

Youngmok Yun [a], Hyun-Chul Kim [b], Sung Yul Shin [b], Junwon Lee [b], Ashish D. Deshpande [a], Changhwan Kim [b,*]

[a] Department of Mechanical Engineering, The University of Texas at Austin, Austin, TX, USA
[b] The Center of Bionics, Korea Institute of Science and Technology, Seoul, South Korea

ARTICLE INFO

Article history:
Accepted 17 September 2013

Keywords:
Gait pattern
Statistics
Gaussian process regression
Stochastic analysis

ABSTRACT

We propose a novel methodology for predicting human gait pattern kinematics based on a statistical and stochastic approach using a method called Gaussian process regression (GPR). We selected 14 body parameters that significantly affect the gait pattern and 14 joint motions that represent gait kinematics. The body parameter and gait kinematics data were recorded from 113 subjects by anthropometric measurements and a motion capture system. We generated a regression model with GPR for gait pattern prediction and built a stochastic function mapping from body parameters to gait kinematics based on the database and GPR, and validated the model with a cross validation method. The function can not only produce trajectories for the joint motions associated with gait kinematics, but can also estimate the associated uncertainties. Our approach results in a novel, low-cost and subject-specific method for predicting gait kinematics with only the subject's body parameters as the necessary input, and also enables a comprehensive understanding of the correlation and uncertainty between body parameters and gait kinematics.

1. Introduction

Many recent attempts to understand and predict human gait pattern utilize model-based optimization techniques. This approach hypothesizes mathematical models and cost functions of the moving bodies, and determines the optimized motion of model. Typical examples of the cost function include total metabolic energy (Anderson and Pandy, 2001) and total muscle effort (Xiang et al., 2011), and the models range from simple link models (Srinivasan and Ruina, 2006; Geyer et al., 2005) to elaborate biomechanical models involving muscles, tendons, varying moment arms and body segments (Anderson and Pandy, 2001; Delp et al., 2007).

While the model-based optimization methods have led to advances in fields such as biomechanics of walking, lower-limb rehabilitation, and biped machine walking (Neptune et al., 2009; Zajac et al., 2003; Bessonnet et al., 2004), there are significant limitations to this approach. The optimization methods need mechanical and biomechanical models for describing the human motion. Assumptions and simplifications are inherent, even in the most advanced models, resulting in limitations of the methods. The choice of cost function affects the results significantly and it is a topic of ongoing scientific debate (Todorov, 2004). Although reasonable choices have been made, it is difficult to capture the complexities of human walking with one specific cost function. Most optimization methods output a deterministic pattern. However, the human movements possess inherent variability and change with time, gender, age, body features and also the emotional state (Blanc et al., 1999; Kerrigan et al., 1998; Cunningham et al., 1982; Escalante et al., 2001; Murray et al., 1964; Samson et al., 2001; Macellari et al., 1999). Moreover, evidence shows that walking may involve randomness (Hausdorff et al., 1995), which is impossible to capture with the deterministic optimization methods, even in a hypothetical case of fully modeled system.

Statistics-based methods provide an alternative approach for gait pattern prediction because these methods are free of any predetermined biomechanical models and cost functions, and can inherently handle the deviations and uncertainty in human walking. Although statistical methods have been employed to investigate the effects of body parameters on the gait pattern, e.g. gender (Blanc et al., 1999; Kerrigan et al., 1998) and age (Cunningham et al., 1982; Escalante et al., 2001), typically, only the effect of a single factor is analyzed and no functional mapping is developed. Over the last two decades, the introduction of motion capture devices in gait pattern research has led to the collection of a large amount of data (Kadaba et al., 1990; Davis et al., 1991). And the

* Corresponding author.
E-mail addresses: yunyoungmok@utexas.edu (Y. Yun), ckim@kist.re.kr (C. Kim).

0021-9290/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jbiomech.2013.09.032

recently developed machine learning algorithms such as neural network (Bishop, 1995), support vector machine (Smola and Schölkopf, 2004), and Gaussian process regression (Rasmussen, 2006) provide novel and computationally powerful statistical tools for processing such a large-scale data.

Our goal is to develop a statistical and stochastic approach for gait kinematics prediction using a large-scale database. We present a method that builds a stochastic function mapping whose input is body parameters and output is the gait pattern kinematics. The developed function mapping is a numerical model, sometimes referred to as a *black-box model*, and it is created through the Gaussian process regression (GPR) algorithm (Rasmussen, 2006). By incorporating the variability in the database in the modeling formulation, GPR can produce a stochastic function whose output is a probability distribution of the predicted gait kinematics rather than a strictly deterministic value.

In addition, we provide a Matlab/Octave toolbox, 'Gait Kinematics Prediction Toolbox' and a database 'KIST Human Gait Pattern Data' for gait kinematics prediction along with a more detailed description of the presented method, as Supplemental Material.

2. Methods

2.1. Data collection with human subjects

A total of 113 healthy subjects (50 males and 63 females) participated in the study. Ethics rules of the Institutional Review Board (IRB) of Korea Institute of Science and Technology (KIST) were followed during data acquisition (IRB Approval number: 2012-006). We chose the subjects so as to maintain a uniform distribution in age between years 20 and 69. For each subject, age and gender were recorded, and 12 body parameters (Fig. 1) were measured using the method described by Chandler et al. (1975). Gait kinematic data were acquired with a motion capture system (Motion Analysis Inc.) with eight cameras, and the Helen Hayes Hospital marker set (15 markers on the lower-body, Kadaba et al., 1990) was adopted. Because we were collecting data with a large number of participants, we decided to attach markers on the subject's clothing rather than on the skin or on a bodysuit. We fixed a velcro belt tightly on the subject's clothed body and then attached markers on the velcro belt (Fig. 2(a)). During the experiment, subjects were asked to walk at a constant speed, 3 km/h, which is a normal walking speed (Ryu et al., 2006) on a treadmill (Fig. 2(b)). When a stable walking speed was reached, motion capture data were recorded for 1 min (approximately 40–60 walking cycles).

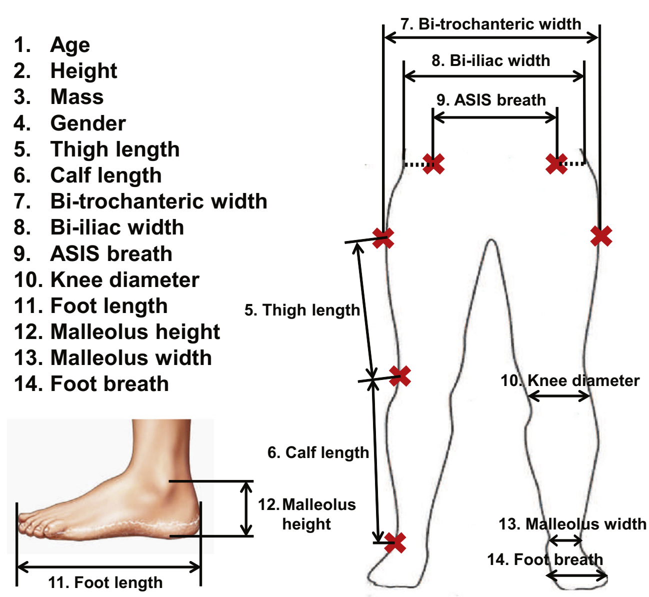

Fig. 1. A total of 14 body parameters, which are easily measurable and significantly affect gait pattern, are selected for this study. Age and gender were recorded through a questionnaire, and other factors were measured by the experimenters with the method described by Chandler et al. (1975).

The 3-D *XYZ* marker position data were converted into human joint space (Kadaba et al., 1990). A marker attached on the pelvis was used as a reference point, and all other relative joint motions were calculated. A total of 14 joint displacements and rotations were calculated from 3-D position data including: pelvis $X/Y/Z$-axis displacement, pelvis rotation, right/left hip adduction/flexion/rotation, right/left knee flexion, and right/left ankle flexion. These 14 kinematic measures are the gait pattern outputs and other minor joint rotations (e.g. knee varus) were not considered in this method.

2.2. Data pre-processing

Raw data from the motion capture device contain 14 joint motions whose time length is about 60 s, including 40–60 gait cycles (Fig. 3(a)). First, for pelvis *XYZ*-axes displacement, the low-frequency offset of the raw data is eliminated by high pass filtering; because the marker on pelvis is used as a reference point, time-variant offset exists. Since the filtered pelvis Y-axis displacement data have the most stable cyclic pattern among all motions (Fig. 3), we used it to determine the cutting points denoting the start and end of one walking cycle (Fig. 3(a)). Second, all raw data of joint motions are divided into several one-period fragments based on the cutting points. This cut in the period was automatically performed by a computer program in Matlab (Mathworks Inc.). The average period is calculated from those different periods (Fig. 3(c)). Third, because the truncated one-period joint motion fragments have varying periods, all one-period joint motions were normalized to have the same period by a time resampling method. Then one average pattern is obtained from the normalized patterns (Fig. 3(d)). The averaged and normalized gait pattern, and averaged period represent one subject's gait pattern. All of the regression procedures are conducted with these normalized gait patterns and averaged periods. The final result of regression is also a normalized gait pattern and a period. Then the normalized gait pattern is stretched back with the predicted period by time resampling.

2.3. Gaussian process regression algorithm for gait pattern prediction

Gait pattern prediction is regarded as a nonlinear regression task and we implemented the Gaussian process regression (GPR) algorithm (Rasmussen, 2006) for generating a functional mapping between the input variables (14 body parameters) and output gait pattern (14 joint motions for gait kinematics and gait period). First, we describe the GPR algorithm to predict joint motions from Sections 2.3.1 and 2.3.3, and then explain the algorithm to predict gait period in Section 2.3.4.

2.3.1. Definition of training set

For the prediction of normalized joint motions, the training input vector $\mathbf{x}$, the training output scalar y, and their sets $\mathbf{X}$ and $\mathbf{y}$ are defined as follows:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_{N\times T}^\top \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1^\top & t_1 \\ \vdots & \vdots \\ \mathbf{b}_1^\top & t_T \\ \mathbf{b}_2^\top & t_1 \\ \vdots & \vdots \\ \mathbf{b}_N^\top & t_T \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_{N\times T} \end{bmatrix} \tag{1}$$

where $\mathbf{b}$ is a vector denoting the subject's body parameter, N is the total number of subjects, t denotes a time index in the normalized time frame, T is the last index of the normalized time frame, and y is a gait pattern value (given as a distance or an angle) corresponding to its $\mathbf{x}$. For example, if a training set is made for ankle flexion pattern, $\mathbf{x}_{(i-1)\times T+j}$ includes the body parameter $\mathbf{b}_i$ of i-th subject and the j-th time index t_j, and corresponding $\mathbf{y}$ is ankle flexion angle for the given condition. This training set is built for all of the 14 joint motions that represent gait kinematics.

2.3.2. Design of Gaussian process model

A Gaussian process is completely defined by its mean function and covariance function (Rasmussen, 2006), as follows:

$$f(\mathbf{x}) \equiv \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{2}$$

where mean function $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$, and covariance function $k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(x))(f(\mathbf{x}') - m(x'))]$.

In our GP model, the mean function and the covariance function are determined as follows:

$$m(\mathbf{x}) = \mathbf{0} \tag{3}$$

$$k(\mathbf{x}, \mathbf{x}') = \nu_0 \exp\left\{ -\frac{(\Delta\mathbf{b})^\top \Lambda_b (\Delta\mathbf{b}) + \lambda_t \Delta t^2}{2} \right\} + \nu_1 \delta(i,j) \tag{4}$$

where $\Delta\mathbf{b}$ is defined as $\mathbf{b} - \mathbf{b}'$ and Δt is defined as $t - t'$. $\delta(i,j)$ is the Kronecker delta function. Λ_b is a diagonal matrix whose elements are $\lambda_{b1} \ldots \lambda_{b14}$. The hyperparameter set is defined as a set of $\nu_0, \nu_1, \lambda_{b1-14}, \lambda_t$, and denoted as Θ. This hyperparameter set determines the characteristics of GP model.

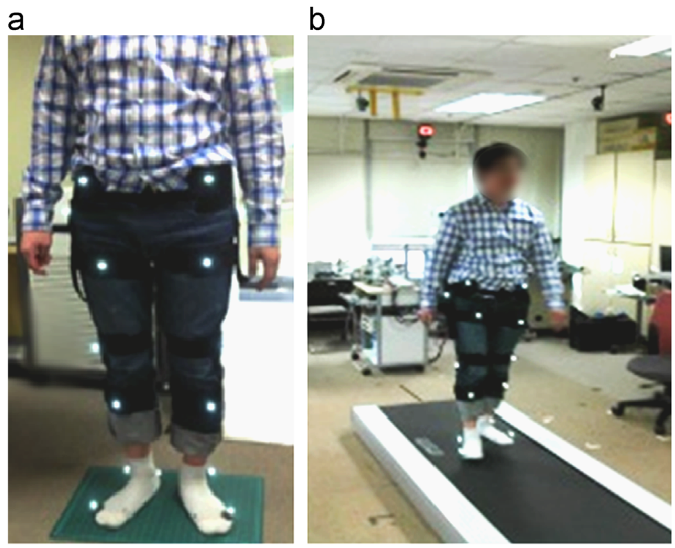

Fig. 2. Motion capture for gait pattern data acquisition. (a) Markers were attached on each subject's body according to the configuration of Helen Hayes Hospital marker set. (b) Subjects walked on a treadmill at a constant speed.

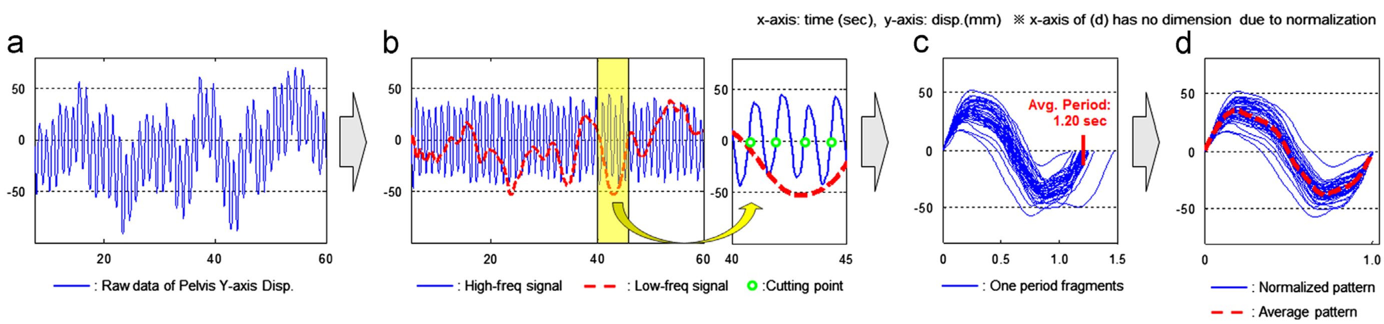

Fig. 3. Data pre-processing procedure. This procedure generates a representative gait pattern for each subject, including an average period and a normalized and averaged gait pattern. (a) Raw joint motion data from the motion capture device were collected for about 60 s. (b) The high-pass filtering step is required only for pelvis *XYZ*-axes displacement data. Raw data for pelvis *XYZ*-axes displacement have time-variant offsets because the marker on the pelvis is a reference point. The low-frequency time-variant offset was eliminated by high-pass filtering. From the high-frequency signal of pelvis *Y*-axis displacement data, we obtained cutting points for period cut; it has the most stable cyclic signal among all joint motions as shown in Fig. 6. (c) All raw data of joints motions are truncated based on the cutting points. Average period is calculated from those different periods. (d) *X*-axis of the truncated fragments are normalized to have the same period by a time resampling method. One average pattern is obtained from the normalized patterns.

With the above configuration, we can predict a probability distribution of joint motion y^* for a given test set $\mathbf{x}^*$ including body parameter $\mathbf{b}^*$ and time t^* as follows:

$$p(y^*|\mathbf{x}^*, \mathbf{X}, \mathbf{y}, \Theta) = \mathcal{N}(\mathbf{k}_*^\top \mathbf{K}^{-1}\mathbf{y}, \kappa - \mathbf{k}_*^\top \mathbf{K}^{-1}\mathbf{k}_*) \tag{5}$$

where $\mathbf{K}$ is a matrix whose elements K_{ij} is a covariance function value of $k(\mathbf{x}_i, \mathbf{x}_j)$. $\mathbf{k}_* = [k(\mathbf{x}^*, \mathbf{x}_1)\ldots k(\mathbf{x}^*, \mathbf{x}_n)]^T$. $\kappa = k(\mathbf{x}^*, \mathbf{x}^*)$.

2.3.3. GP model optimization

The newly devised GP model for gait pattern prediction is defined by the mean function (3) and the covariance function (4), and their characteristics are adjusted by the hyperparameter set Θ. By selecting the hyperparameter set, the optimal GP model which most well represents all of the training set was selected. The optimization of the hyperparameter set was performed by maximizing the likelihood among training dataset, or $p(\mathbf{y}|\mathbf{X}, \Theta)$. Practically, the optimization was performed by maximizing log-likelihood as shown in (6). In this study, the optimization was performed by a line search algorithm (Press et al., 2009):

$$\log\ p(\mathbf{y}|\mathbf{X}, \Theta) = -\frac{1}{2}\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y} - \frac{1}{2}\ \log|\mathbf{K}| - \frac{NT}{2}\ \log\ 2\pi \tag{6}$$

2.3.4. Gait period prediction

Compared with the prediction of joint motion, the prediction of the gait period is relatively simple because its training set does not include time index. For the prediction of the gait period, the training data and the covariance function are set as follows:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_N^\top \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1^\top \\ \vdots \\ \mathbf{b}_N^\top \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \tag{7}$$

$$k(\mathbf{x}, \mathbf{x}') = v_0 \exp\left\{ -\frac{(\Delta\mathbf{b})^\top \Lambda_b (\Delta\mathbf{b})}{2} \right\} + v_1 \delta(i, j) \tag{8}$$

Here, y_i is *i*-th subject's average period. All the other GPR procedures are performed in the previously described procedure.

2.4. Validation process

To validate performance of the proposed prediction algorithm, a cross-validation method (also called the Leave-one-out cross-validation, Kohavi et al., 1995, or the Jackknife method Miller, 1974) was adopted by regarding one test subject's gait pattern data as unknown and then using the other subjects' data as the regression training set for generating a functional mapping. After predicting the gait pattern for the test subject using his/her body parameters and the developed functional mapping, the actual and predicted gait patterns are quantitatively compared. These steps are repeated by considering each subject as a test subject for validating the proposed methodology.

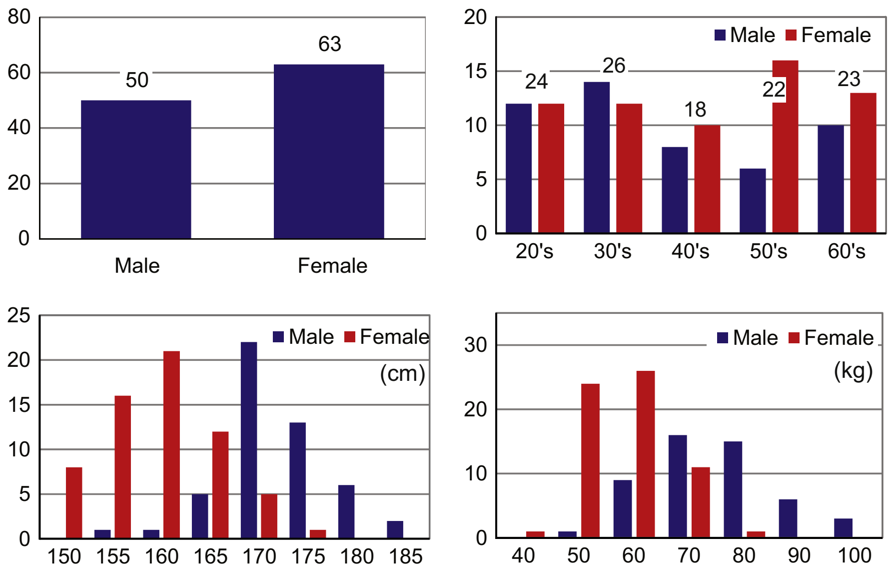

Fig. 4. A total of 113 subjects participated in the experiments. (a) Gender distribution, (b) age distribution, (c) height distribution, and (d) weight distribution.

Table 1
Mean and standard deviation of 113 experiment participants' body parameters. The mean and standard variation of gender is not calculated due to its discrete characteristic.

Body parameters	Mean	Std.
Age (years)	44.27	14.92
Height (cm)	164.8	8.323
Mass (kg)	65.74	12.48
Thigh length (cm)	35.37	3.151
Calf length (cm)	39.16	3.131
Bi-trochanteric width (cm)	33.29	1.753
Bi-iliac width (cm)	30.71	2.125
ASIS breath (cm)	25.71	2.300
Knee diameter (cm)	10.33	0.898
Foot length (cm)	23.85	1.501
Malleolus height (cm)	6.696	0.802
Malleolus (cm)	6.735	0.563
Foot breath (cm)	9.412	0.724

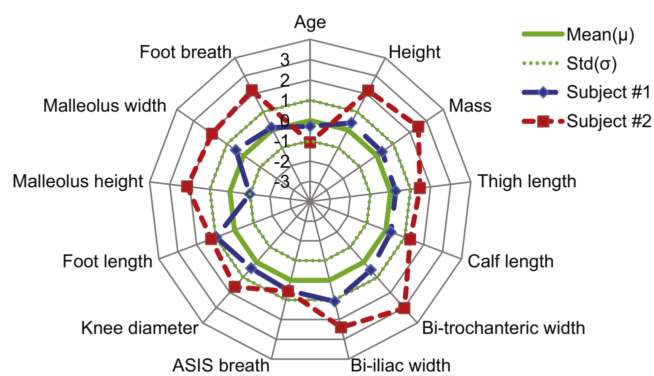

Fig. 5. The distribution of body parameters of a female subject S1 (blue) and a male subject S2 (red). The green line represents mean value of all subjects' body parameters and green dot shows its standard deviation. All distributions are normalized, thus 0 indicates the mean (μ) and 1 means one standard deviation (σ) value. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

3. Results

The distribution of subjects shows comparable numbers of males and females, uniform distribution in age from 20-year-old to 69-year-old, and normal distributions for height and weight (Fig. 4). Table 1 lists the mean values and standard deviations for all the body parameters for the subjects (except for gender).

We have built a functional map between body parameters and gait kinematics with data from 113 subjects and the function predicts not only the trajectories of an arbitrary person's 14 joint motions representing gait patterns but also the uncertainty associated with each of the motions. We illustrate the prediction methodology with results from two subjects: S1 and S2. S1's body parameters are close to the average of all the subjects and S2's body parameters far from the average values (Fig. 5 and Table 1). Figs. 6 and 7 show the predicted gait motions for the two subjects.

Through the cross-validation method, the mean and standard deviation of predicted gait patterns for all subjects were calculated. We present the results for subjects S1 and S2, and also the overall results (Table 2). The quantitative measures for cross-validation include the *mean error*, which indicates the difference between the actual and the predicted mean values, the average values of *predicted standard deviation*, and a term called *auto mean error* which is the difference between the actual and the average of the trajectories over all the subjects. Mathematical definitions are described in Table 2.

4. Discussion

The main contribution of this paper is a novel statistical methodology for predicting human gait pattern kinematics. A stochastic functional mapping is created based on a large database and the GPR algorithm. The presented approach is a low cost method for predicting gait kinematics because it only needs 14 easily measured body parameters. The output of the function is not a strict value but a probabilistic distribution providing the trajectory of gait pattern and its uncertainty. The method enables a comprehensive understanding of the correlations and associated uncertainty between body parameters and gait kinematics. Previous statistical methods thus far have analyzed correlations between a particular gait feature (e.g. stride length) and body parameter (e.g. height), but this new approach allows for the simultaneous analysis of multiple factors. This may lead to better modeling and analysis of human walking.

One valuable use of the method is the prediction of a normal gait pattern for a subject with physical or neurological deficiencies (e.g.

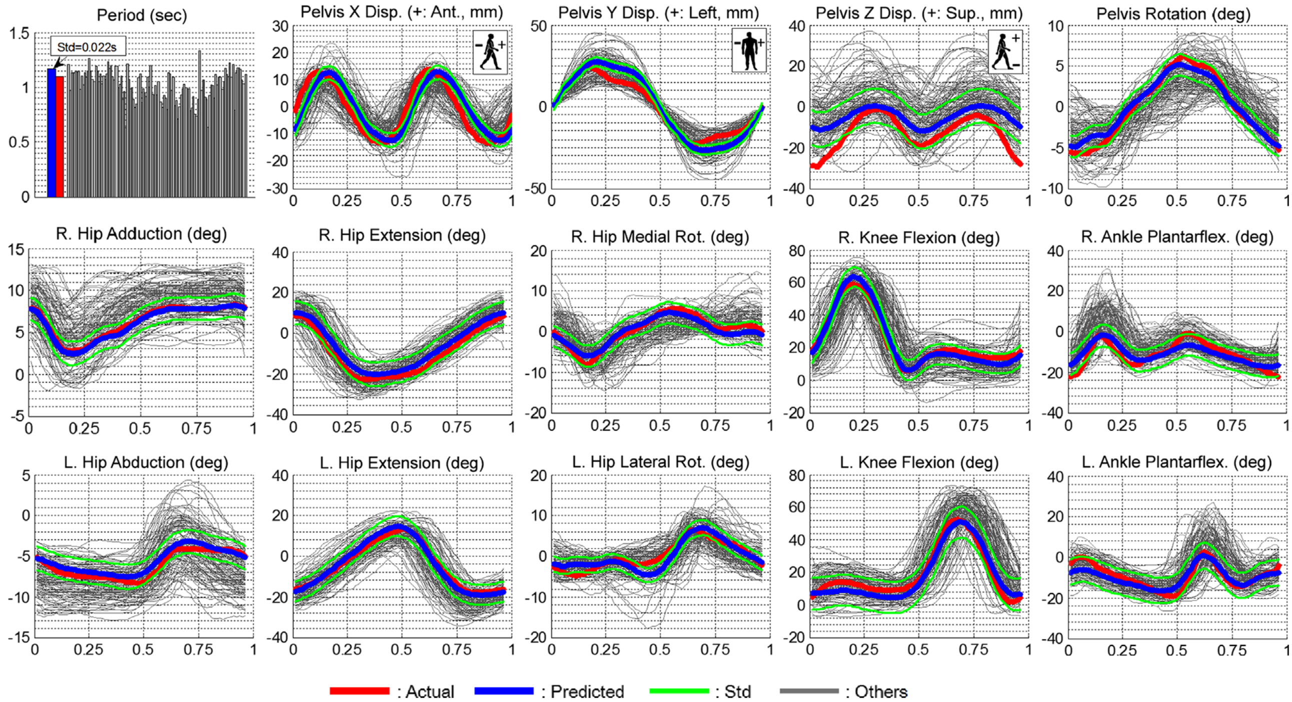

Fig. 6. Gait pattern prediction result using GPR with subject S1's body parameters, which are close to the average values of all subjects as shown in Fig. 5. *X*-axis indicates normalized time in all the figures except the top left. The suggested algorithm provides both the mean and the standard deviation of predicted joint motion pattern and gait period. Gait patterns for all other subjects are drawn with gray lines, and these were used as a training output.

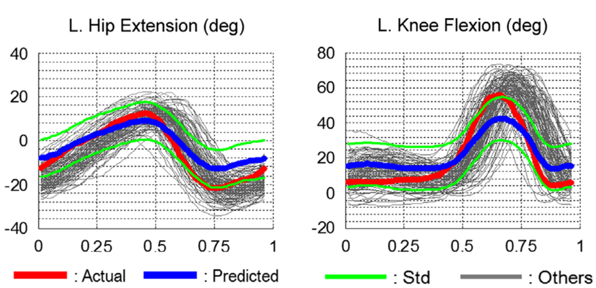

Fig. 7. Predicted joint motion results for S2's left hip extension and left knee flexion. *X*-axis indicates normalized time line. S2's body parameters are far from the average of training set's as shown in Fig. 5, and this results in large uncertainty of predicted joint motions as shown in the wide width of green lines. For the full prediction result, refer to the Supplementary Material.(For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

amputation or stroke). In most rehabilitation and prosthetics implementations, the target gait pattern is an *average* motion. The presented method predicts gait pattern specific to a subject which is more accurate than average of motions from all the subjects (Table 2). Knowing the subject-specific healthy gait pattern will allow therapists to better plan and monitor rehabilitation, and allow designers to customize prostheses. Other potential applications include the generation of reference trajectories for the control of a humanoid robot or an animated character in order to generate a walking pattern that closely resembles a particular subject's gait.

4.1. *Mean error and uncertainty of predicted gait pattern*

The proposed algorithm provides a method for predicting probability distributions corresponding to a test subject's joint motions representing his/her gait pattern (Fig. 6 and Table 2). For all subjects, mean error (footnote c in Table 2) is smaller than auto mean error (footnote b in Table 2), which confirms that the prediction algorithm is effective.

The mean value of the probability distribution is the predicted joint motion trajectory for the subject, and the standard deviation of the distribution indicates how precisely we can predict the trajectory. The accuracy of prediction and the size of standard deviation for a given subject are affected by two factors: the uncertainty of training data itself, and the distribution of a test subject's body parameters compared to the distribution of training subjects' body parameters in the database. A joint motion with high repeatability over time and over different subjects in the database is predicted with small mean error and low standard deviation, for example pelvis *Y* displacement for subject S1 (Fig. 6), while mean errors are larger in the prediction of motion with low repeatability over subjects, for example pelvis *Z* displacement for subject S1 (Fig. 6). Comparison of prediction results obtained for the subjects S1 and S2 shows the effects of body parameter distribution on the accuracy of prediction. The values for the body parameters for the subject S1 are close to the mean values of database (Fig. 5) so the function predicts subject S1's motion trajectories with small errors and low standard deviation (Fig. 6 and Table 2); whereas the body parameters for subject S2 are much larger than the mean values of the database (Fig. 5) and motion trajectories for this subject are predicted with large errors and wide standard deviation (Fig. 7 and Table 2).

The predicted uncertainty, or predicted standard deviation, is a valuable feature for our statistical regression method because it provides the *confidence level* of the prediction. From the predicted standard deviation, we can guess the size of mean error, and actually the mean error and predicted standard deviation have the same trend (Table 2). In statistical approaches, prediction error, which comes from various sources including a sparse database, measurement error in the data collection, and the effect of uncounted body parameters, is inevitable. However, if we know the level of confidence for the prediction, researchers can assign appropriate actions to the prediction results. Also the confidence

Table 2
The cross validation results for gait pattern prediction.

Predicted feature	S1		S2		All subjects		
	Mean err.[a]	Predicted std[b]	Mean err.[a]	Predicted std.[b]	Mean err.[c]	Predicted std.[d]	Auto mean err.[e]
Period (s)	0.055	0.022	0.017	0.033	0.068	0.041	0.137
Pelvis *X* disp. (mm)	2.300	2.428	6.037	6.857	3.462	3.355	5.045
Pelvis *Y* disp. (mm)	3.741	1.525	6.264	7.572	3.583	3.934	5.046
Pelvis *Z* disp. (mm)	6.438	7.334	8.203	8.696	6.651	3.911	9.725
Pelvis rot. (deg)	0.886	1.201	1.369	2.078	1.265	1.513	1.894
R. Hip add. (deg)	0.251	1.355	2.788	1.974	1.492	0.821	2.240
R. Hip ext. (deg)	2.129	1.956	6.141	8.151	3.995	2.427	5.491
R. Hip M. rot. (deg)	1.126	2.620	3.163	3.255	2.026	1.750	2.835
R. Knee flex. (deg)	2.573	4.970	10.170	12.670	7.055	4.232	9.520
R. Ankle P.flex. (deg)	3.462	5.333	5.743	6.094	4.295	2.755	5.567
L. Hip abd. (deg)	0.712	1.726	1.369	1.870	1.512	0.796	2.238
L. Hip ext. (deg)	1.507	4.765	5.209	7.287	3.981	2.346	5.627
L. Hip L. rot. (deg)	1.553	1.832	1.861	2.865	1.735	1.747	2.810
L. Knee flex. (deg)	3.918	4.978	8.402	13.750	6.952	4.238	9.372
L. Ankle P. flex. (deg)	1.759	5.210	4.517	6.444	4.203	2.631	5.570

[a] The mean error for S1 and S2 is defined as $e(s,j)=(1/T)\sum_{t=1}^{T}|y^*(s,j,t)-y(s,j,t)|$.

[b] The predicted standard deviation is defined as $\sigma(s,j)=(1/T)\sum_{t=1}^{T}\sigma(s,j,t)$. s is the subject index, j is the joint index, t is the time index, T is the max time index. y^* is the predicted mean value, y is the actual value, and σ is the predicted standard deviation.

[c] The mean error for all subjects is defined as $e(j)=(1/S)\sum_{s=1}^{S}e(s,j)$.

[d] The predicted standard deviation for all subjects is defined by the average value of predicted standard deviations, or $\sigma(j)=(1/S)\sum_{s=1}^{S}\sigma(s,j)$. S is the total number of subjects.

[e] The auto mean error is defined by the deviation of training trajectory itself. In other words, it is obtained from the difference between the training trajectories and the mean of themselves. $d(j)=(1/T)\sum_{t=1}^{T}[(1/S)\sum_{s=1}^{S}|y(s,j,t)-y(j,t)|]$ where $y(j,t)=(1/S)\sum_{s=1}^{S}y(s,j,t)$.

level may aid experimenters in deciding if the collected database for training is sufficiently rich for their application.

4.2. Body parameters and gait kinematics

A large number of factors influence gait kinematics for a subject, including age, gender, height, weight, body fat, muscle strength and even psychological condition. It is prohibitively difficult and also undesirable to include all the factors in the modeling, analysis and prediction of gait patterns with a statistical method. While ignoring key factors may limit the effectiveness of the prediction methods, including too many factors will lead to very high dimension of input for the training data leading to degradation of the performance of the regression algorithm. Thus, the number and type of input variables should be chosen carefully. We chose 14 representative human features which significantly affect the gait kinematics (Vaughan et al., 1992; Murray, 1967; Hanavan, 1964) and which are easy to measure (Chandler et al., 1975), thus allowing for data collection with a large number of subjects. The effects of the other factors are encapsulated by uncertainties in the stochastic formulation of GPR.

Numerous previous research approaches have investigated the relationship between the body parameters and gait parameters (not a trajectory motion) with statistical methods. They analyzed that the effects of one or two factors at a time and typically simple (linear) dependency is assumed (Blanc et al., 1999; Kerrigan et al., 1998; Cunningham et al., 1982; Escalante et al., 2001). However, the relationships between gait pattern and body parameters are nonlinear and correlated with one another. For example, female subjects generally walk with shorter strides than male subjects, but when the stride length is normalized by height, females shows a similar or slightly larger stride length (Escalante et al., 2001; Kerrigan et al., 1998), thus suggesting that simultaneously analyzing effects of multiple human parameters may lead to more insightful understanding. The proposed method allows, for the first time, a general mapping function over body parameters and gait pattern. By controlling the input vector to this function, the effects of multiple factors can be analyzed simultaneously.

4.3. Regression algorithm

A number of methods exist for nonlinear regression such as least squares and its variants (Bates and Watts, 2007), neural networks (Bishop, 1995), and support vector regression (Smola and Schölkopf, 2004). We chose to employ the Gaussian process regression (Rasmussen, 2006) for the following reasons: First, GPR is a powerful nonlinear regression algorithm for a database whose input (here, body parameters) has high dimension and the elements of the input are highly correlated. The GPR algorithm only counts the Euclidean distance between body parameter sets in the scaled input space, $(\Delta \mathbf{b})^{\top}\Lambda_b(\Delta \mathbf{b})$, as shown in (4). Even though the input vector has high dimension, if their Euclidean distances are close, its prediction is reliable. Body parameters are highly correlated, and thus their distribution is neither even nor sparse in the input space, instead it is located in a narrow region called the reduced-order manifold or latent variable space. Therefore, the GPR algorithm can reliably predict an arbitrary subject's gait kinematics by comparing his/her body parameters with the training subjects' body parameters lying on the narrow region. Second, GPR inherently assumes the uncertainty of the training set and produces a stochastic function. Compared with other deterministic algorithms, GPR is robust to the errors in training set and we can evaluate its reliability from the predicted standard deviation value. Third, GPR has a principled way for the selection of complexity of models. Once the GP is defined with a mean and a covariance function, the complexity of the model is determined by the database. This property is an important advantage of GPR compared with other traditional nonlinear regression methods, such as least squares methods, neural networks, and support vector regression.

4.4. Limitations

While this completely statistics-based approach for predicting gait kinematics, with no analytical models, provides broader freedom for modeling human walking, some advantages gained from insightful models are missing in the presented approach. Also, at this point, only the kinematics of gait patterns are analyzed. The limitations can be addressed by harnessing the power of numerical tools and analytical models. One method is to implement biomechanical models as kinematic and dynamic constraints into the GPR algorithm (Wang et al., 2006). Another approach would be to develop analytical models and then use GPR to modify these models. In this study we only chose input parameters representing kinematics of gait. To incorporate gait biomechanics and dynamics, we will choose biomechanical and kinetic parameters such as link inertias and joint stiffness. In this paper, we have only analyzed walking at one speed on leveled ground. Analysis of effect of walking speed on gait kinematics is part of our ongoing work. The accuracy and reliability of the presented method depends on the size of database, which is true for all statistical methods.

Conflict of interest statement

None of the authors has any financial or personal relationships with other people or organizations that could inappropriately influence (bias) his work.

Acknowledgements

This work was supported partly by Korea Institute of Science and Technology (Project no. 2V01500). The authors thank Sang Hyup Lee and a number of other research colleagues at the Korea Institute of Science and Technology (KIST) for their help with the collection of human subject data.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jbiomech.2013.09.032.

References

Anderson, F., Pandy, M., 2001. Dynamic optimization of human walking. J. Biomech. Eng. 123, 381.
Bates, D.M., Watts, D.G., 2007. Nonlinear Regression Analysis and Its Applications. Wiley-Interscience.
Bessonnet, G., Chessé, S., Sardain, P., 2004. Optimal gait synthesis of a seven-link planar biped. Int. J. Robot. Res. 23 (10–11), 1059–1073.
Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, Inc., New York, NY, USA.
Blanc, Y., Balmer, C., Landis, T., Vingerhoets, F., 1999. Temporal parameters and patterns of the foot roll over during walking: normative data for healthy adults. Gait Posture 10 (2), 97–108.
Chandler, R., Clauser, C., McConville, J., Reynolds, H., Young, J., 1975. Investigation of Inertial Properties of the Human Body. Technical Report, DTIC Document.
Cunningham, D., Rechnitzer, P., Pearce, M., Donner, A., 1982. Determinants of self-selected walking pace across ages 19 to 66. J. Gerontol. 37 (5), 560–564.
Davis, R., Ounpuu, S., Tyburski, D., Gage, J., 1991. A gait analysis data collection and reduction technique. Hum. Mov. Sci. 10 (5), 575–587.
Delp, S., Anderson, F., Arnold, A., Loan, P., Habib, A., John, C., Guendelman, E., Thelen, D., 2007. Opensim: open-source software to create and analyze dynamic simulations of movement. IEEE Trans. Biomed. Eng. 54 (11), 1940–1950.
Escalante, A., Lichtenstein, M., Hazuda, H., 2001. Walking velocity in aged persons: its association with lower extremity joint range of motion. Arthritis Care Res. 45 (3), 287–294.
Geyer, H., Seyfarth, A., Blickhan, R., 2005. Spring-mass running: simple approximate solution and application to gait stability. J. Theor. Biol. 232 (3), 315–328.
Hanavan, Jr., E., 1964. A Mathematical Model of the Human Body. Technical Report, DTIC Document.
Hausdorff, J., Peng, C., Ladin, Z., Wei, J., Goldberger, A., 1995. Is walking a random walk? evidence for long-range correlations in stride interval of human gait. J. Appl. Physiol. 78 (1), 349–358.
Kadaba, M., Ramakrishnan, H., Wootten, M., 1990. Measurement of lower extremity kinematics during level walking. J. Orthop. Res. 8 (3), 383–392.
Kerrigan, D., Todd, M., Della Croce, U., et al., 1998. Gender differences in joint biomechanics during walking: normative study in young adults. Am. J. Phys. Med. Rehabil. Assoc. Acad. Physiatr. 77 (1), 2.
Kohavi, R., et al., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: International joint Conference on Artificial Intelligence, vol.14. Lawrence Erlbaum Associates Ltd, pp. 1137–1145.
Macellari, V., Giacomozzi, C., Saggini, R., 1999. Spatial-temporal parameters of gait: reference data and a statistical method for normality assessment. Gait Posture 10 (2), 171–181.
Miller, R.G., 1974. The jackknife-a review. Biometrika 61 (1), 1–15.
Murray, M., 1967. Gait as a total pattern of movement: including a bibliography on gait. Am. J. Phys. Med. Rehabil. 46 (1), 290.
Murray, M., Drought, A., Kory, R., 1964. Walking patterns of normal men. J. Bone Joint Surg. (Am.) 46 (2), 335–360.
Neptune, R., Clark, D., Kautz, S., 2009. Modular control of human walking: a simulation study. J. Biomech. 42 (9), 1282–1287.
Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2009. Numerical Recipes in C+: The Art of Scientific Computing, vol. 994. Cambridge University Press, Cambridge.
Rasmussen, C.E., 2006. Gaussian Processes for Machine Learning.
Ryu, T., Soon Choi, H., Choi, H., Chung, M., 2006. A comparison of gait characteristics between Korean and western people for establishing Korean gait reference data. Int. J. Ind. Ergon. 36 (12), 1023–1030.
Samson, M., Crowe, A., De Vreede, P., Dessens, J., Duursma, S., Verhaar, H., et al., 2001. Differences in gait parameters at a preferred walking speed in healthy subjects due to age, height and body weight. Aging (Milan, Italy) 13 (1), 16.
Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14 (August (3)), 199–222.
Srinivasan, M., Ruina, A., 2006. Computer optimization of a minimal biped model discovers walking and running. Nature 439 (7072), 72–75.
Todorov, E., 2004. Optimality principles in sensorimotor control. Nat. Neurosci. 7 (9), 907–915.
Vaughan, C., Davis, B., O'connor, J., 1992. Dynamics of Human Gait. Human Kinetics Publishers Champaign, Illinois.
Wang, J., Fleet, D., Hertzmann, A., 2006. Gaussian process dynamical models. Adv. Neural Inf. Process. Syst. 18, 1441.
Xiang, Y., Arora, J., Abdel-Malek, K., 2011. Optimization-based prediction of asymmetric human gait. J. Biomech. 44 (4), 683–693.
Zajac, F., Neptune, R., Kautz, S., 2003. Biomechanics and muscle coordination of human walking: part ii: lessons from dynamical simulations and clinical implications. Gait Posture 17 (1), 1–17.