Published in final edited form as:

IEEE Trans Biomed Eng. 2011 October ; 58(10): 2804–2815. doi:10.1109/TBME.2011.2159840.

An Experimental Study in Determining Energy Expenditure from Treadmill Walking using Hip-Worn Inertial Sensors

Harshvardhan Vathsangam,

Department of Computer Science, University of Southern California, Los Angeles, CA 90089

Adar Emken,

Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089

E. Todd Schroeder,

Division of Biokinesiology and Physical Therapy, University of Southern California, Los Angeles, CA 90089

Donna Spruijt-Metz, and

Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089

Gaurav S. Sukhatme

Department of Computer Science, University of Southern California, Los Angeles, CA 90089

Abstract

This paper describes an experimental study in estimating energy expenditure from treadmill walking using a single hip-mounted triaxial inertial sensor comprised of a triaxial accelerometer and a triaxial gyroscope. Typical physical activity characterization using accelerometer generated counts suffers from two drawbacks - imprecision (due to proprietary counts) and incompleteness (due to incomplete movement description). We address these problems in the context of steady state walking by directly estimating energy expenditure with data from a hip-mounted inertial sensor. We represent the cyclic nature of walking with a Fourier transform of sensor streams and show how one can map this representation to energy expenditure (as measured by $\dot{V}O_2$ consumption, mL/min) using three regression techniques - Least Squares Regression (LSR), Bayesian Linear Regression (BLR) and Gaussian Process Regression (GPR). We perform a comparative analysis of the accuracy of sensor streams in predicting energy expenditure (measured by RMS prediction accuracy). Triaxial information is more accurate than uniaxial information. LSR based approaches are prone to outlier sensitivity and overfitting. Gyroscopic information showed equivalent if not better prediction accuracy as compared to accelerometers. Combining accelerometer and gyroscopic information provided better accuracy than using either sensor alone. We also analyze the best algorithmic approach among linear and nonlinear methods as measured by RMS prediction accuracy and run time. Nonlinear regression methods showed better prediction accuracy but required an order of magnitude of run time. This paper emphasizes the role of probabilistic techniques in conjunction with joint modeling of triaxial accelerations and rotational rates to improve energy expenditure prediction for steady-state treadmill walking.

Keywords

Accelerometer; Energy expenditure; Gyroscope; Treadmill walking

I. Introduction

Physical inactivity (lack of physical activity) is the fourth leading risk factor for global mortality (6% of global deaths) and is estimated to be the main cause for approximately 21–

25% of breast and colon cancers, 27% of diabetes and approximately 30% of ischaemic heart disease burden [1]. Closely related to physical inactivity is the problem of obesity. 45% of the United States population is expected to be obese by 2020 [2]. Obesity brings with it reduced life expectancy and loss of quality. Measures that promote a more active lifestyle are therefore the need of the hour. Regular physical activity is known to reduce obesity, risks for cardiovascular disease, type 2 diabetes, and several forms of cancer [3]. In the quest to promote healthier lifestyles, the availability of portable technological solutions provides users and clinicians with objective measures of activity intensity that can be used in feedback and interventions. The challenge is further heightened by requiring that these tools be available in real time.

In recent years, considerable research has been directed towards the detection and classification of physical activity patterns from body mounted inertial sensors [4]. Inertial sensors capture movement either by measuring body accelerations (accelerometers) or rotational rates (gyroscopes). Owing to their small size, low cost, increasingly high precision, low power consumption and portability, inertial sensors have evolved into an attractive option for deriving relevant physiological quantities from human movement [5]. Inertial sensors capture movement of the body to which they are attached. By virtue of being able to detect, classify and measure intensity of movement [6], these sensors provide a cheap and portable way to quantify movement.

The problem of energy expenditure prediction

Walking is an easy and common type of activity that can be used to maintain an active lifestyle [7]. An objective measure of the intensity of walk is energy expended. Characterizing energy expenditure due to walking could be a vital tool to assess physical activity levels and determine the effectiveness of interventions. Walking is a dynamic activity and inertial sensors by definition capture movement. An immediate question to ask is whether raw data from an inertial sensor can be used to predict energy expenditure. Specifically, can properties of movement be exploited to provide intuitive models of expenditure prediction? Also, is it possible to exploit these properties in combination with sophisticated modeling techniques to provide better energy prediction than previously available? Finally, it remains to be explored whether certain kinds of kinematic sensors are better than others at predicting energy expenditure.

The domain

In this paper, we address the problem of estimating energy expenditure using body mounted inertial sensors for on one particular activity: treadmill walking. We use inertial sensor data from a triaxial accelerometer and triaxial gyroscope mounted on the right iliac crest to describe movement and show how one can relate properties of this movement viz. periodicity to energy expenditure. We treat the functional mapping of these frequency characteristics to energy expenditure as a regression problem and show how one can use these maps to build regression models and compare the predictive capability of kinematic sensors.

Contributions of this paper

Our approach to estimating energy expenditure from walking involves developing a probabilistic map from movement features to calories burned. We describe two experimental studies. The first is a comparative analysis of different feature combinations of sensor streams from triaxial accelerometers and gyroscopes to find the most accurate representative space to predict energy expenditure for treadmill walking (measured by RMS prediction accuracy). Our analysis showed that when only one sensor is used, triaxial information from either sensor is advantageous to using information any single axis. Our study also

demonstrated the inadequacy of using simple least-squared based linear regression approaches in predicting energy expenditure due to the effects of outlier sensitivity and overfitting. Using gyroscopic information shows equivalent if not better prediction accuracy as compared to accelerometers due to readings being free from gravitational bias. Combining accelerometer and gyroscopic information provides better accuracy than using either sensor alone. The second study is an analysis of the best algorithmic approach among linear and nonlinear methods as measured by RMS prediction accuracy and run time. Our study showed that nonlinear regression methods show better prediction accuracy but require an order of magnitude of run time. Therefore, in resource-constrained systems, it would be advisable to trade-off accuracy for linear methods with lesser run time. In doing so we found the optimal model for personalized energy expenditure prediction using hip-worn inertial sensors and probabilistic mapping techniques as applied to treadmill walking. A preliminary version of this work has been reported [8].

This paper is organized as follows. We discuss related work in the following section. In Section III, the relation between periodicity of walk and energy expenditure is established from a qualitative perspective. In Section IV, methods of mapping this periodicity to energy expenditure are described from a theoretical perspective. In Section V, data collection and training methodologies are described. Section VI provides a comparative analysis of using different sensor streams and algorithmic approaches in predicting energy expenditure from inertial data. Concluding remarks are in Section VII.

II. Related Work

Much of the research involving using inertial sensors to calculate energy expenditure for daily activities has focused on the utility of accelerometers alone [9], [10]. Accelerometers provide the means to examine the intensity, frequency, and duration of physical activity objectively and independent of location. There exists a wealth of literature in using commercial accelerometers to predict energy expenditure in a variety of settings [11]. A major drawback in using commercial accelerometers is imprecision. Imprecision is because commercial accelerometers use proprietary methods to convert linear accelerations into epoch-based “counts” that are converted to caloric expenditure [12], [13], [14], [15]. The usage of counts does not have either a meaningful or physically interpretable intuition [16]. Also, many accelerometry based techniques involve fitting regression equations that map counts to energy expenditure [17], [18]. Standard linear regression does not mirror the significance of differing amounts of data available to derive model parameters. Single variable linear regression models are limited in that regression mapping to energy can be made richer by considering multi-dimensional features simultaneously [11]. An alternative approach is to define a joint probabilistic map from representative features to the target quantity. By virtue of being defined as a probability distribution, a predicted value will have a characteristic mean and variance instead of a single point estimate. This provides not only the most likely estimate but also the confidence of prediction. Using a confidence measure that is derived from a training dataset helps assessing the quality of a prediction.

An alternative approach to characterizing human motion involves pattern recognition techniques that extract meaningful properties or features from raw movement data and map these properties to calories burnt [19]. These include neural networks [20], probabilistic linear regression [8] and piecewise regression [21]. Using such techniques, it is possible to “learn” a personalized model for each user from data collected. Access to raw data allows the researcher to explore the physical intuition behind movement and use features that explicitly mirror the quantity in question.

Using accelerometry-only techniques suffers from a second limitation, incompleteness. The assumption behind using accelerometry for physical activity monitoring is that data from an accelerometer represents body movement [22]. However, rigid body movement consists of both accelerations and rotations [23] and rotational data cannot be completely separated from translational data using a single triaxial accelerometer [24]. Current count-based accelerometry completely ignores rotational rates. Combining accelerometry and rotational rate measurements through gyroscopes would thus be a valuable tool in completely characterizing movement. Gyroscopes are not influenced by gravity acceleration and are more displacement tolerant than accelerometers because for a given body segment movement, a gyroscope provides the same readings irrespective of position as long as the axis of placement is parallel to the measured axis [4]. The introduction of lowcost, single-chip triaxial gyroscopic sensors [25] has provided the possibility of using gyroscopes as alternatives to or in combination with accelerometers for activity characterization.

III. Analysis of Treadmill Walking Inertial Data

A. Periodicity of Human Walk

Steady state walking is cyclic in nature [26]. Research has examined the gait cycles in walking and the associated periodic limb movements [23]. In our approach, this inherent periodicity in walking was captured with inertial sensor data from the iliac crest on the right hip. Movement signals corresponded directly to the accelerations and rotational rates of the hip as felt by the sensor in its local frame of reference. Fig. 1 shows sample inertial data from treadmill walking collected over 10 seconds when walking at a speed of 2.5 mph. Regular periodic patterns were observed in steady-state. Similar patterns were observed for other speeds.

B. Variation of periodicity with speed

The periodicity of walking signals was examined by computing their Fourier transforms. Figure 2 illustrates Fourier transforms taken from two 10 second steady state walking samples at 2.5 mph and 3.5 mph for the X-axis acceleration streams. The Fourier transform showed clear peaks indicating distinct periodic components for the original signals. The peaks occurred at the same frequencies for all other sensor streams as well. Furthermore, the location of these peaks was a function of walking speed. The dominant peak for walking at 3.5 mph occurs at a higher frequency than the corresponding peak for walking at 2.5 mph. Walking at higher speeds requires a smaller time period per gait cycle implying higher frequency components. Fig. 3 illustrates the variation of the Fourier transform peaks with different speed. Distinct peaks exist for each speed and peaks shift to higher frequencies with increasing speed. The relationship between energy expenditure and walking speed has often been modeled as one where walking at higher speeds requires higher energy expenditure [27], [28], [29]. The fact that each speed exhibits a characteristic frequency spectrum and that there exists a map between speed and energy expenditure suggests that one can track calories burnt from treadmill walking using the frequency spectrum as a representative feature space. In our approach, rather than modeling this map based on biomechanical considerations, we learned a customized, probabilistic map based on training data.

IV. Mapping Walking Data to Energy Expenditure

Given the representation of treadmill walking using features described in Sec. III, we focused on the problem of deriving a mapping from these features to energy consumed as measured by $\dot{V}O_2$ consumption (mL/min). This was framed as a regression problem.

A. Problem Formulation

Consider a D-dimensional *input variable* $\mathbf{x} \in \mathbb{R}^D$ of which there are specific data points $\{x_n\}_{n=1}^N$. The goal of regression is to predict the value of one or more continuous *target variables* $\mathbf{t}$ of which there are corresponding observed values $\{t_n\}_{n=1}^N$ that are related to the input variables by a function $f(x_n)$. Regression aims at finding the “best-fit” function f so that predictions from this function are as close to the target value t_n as possible. This section examines three candidate algorithms to find this map. We provide descriptions of each algorithm for the case of an arbitrary D-dimensional input variable and one dimensional target variable and discuss the relative merits and demerits of each to motivate our choice.

B. Least Squares Linear Regression

Least Squares Regression (LSR) [30] models regression as a linear combination of input variables. Specifically, for an input data point $\mathbf{x}_n$, we have

$$\begin{aligned} t_n &= w_0 + w_1 \phi_1(x_n) + \dots w_{M-1} \phi_{M-1}(x_n) + \epsilon \\ \text{with } \epsilon &\sim \mathcal{N}(0, \beta^{-1} \mathbf{I}) \\ \Rightarrow t_n &= \mathbf{w}^T \phi(x_n) + \epsilon, \epsilon \sim \mathcal{N}(0, \beta^{-1} \mathbf{I}) \end{aligned} \quad (1)$$

where ϵ is a noise parameter, $\phi = (\phi_0, \dots, \phi_{M-1})^T$ is the derived function space consisting of fixed nonlinear functions of the input variables of dimension $M - 1$ and $\mathbf{w} = (w_0, \dots, w_{M-1})^T$ are the weights. This model is linear in ϕ . This allows the usage of feature functions $\{\phi_j(x_n)\}$ derived from input variables x_n . Equation 1 describes a mapping from feature space $\{\phi_j(\mathbf{x}_n)\}$ to the output values t_n . Allowing a probabilistic interpretation, we have

$$p(t_n | x_n, \mathbf{w}, \beta) = \mathcal{N}(t_n; \mathbf{w}^T \phi(x_n), \beta^{-1}) \quad (2)$$

We define an optimal fitting function as one that maximizes the likelihood

$p(\mathbf{t} | \mathbf{w}_T \mathbf{X}) = \prod_{n=1}^N p(t_n | x_n, \mathbf{w}, \beta)$. This is equivalent to finding the optimal $\mathbf{w}$ that would minimize the expected square-loss $E_D \{(t_n - f(x_n, \mathbf{w}))^2\}$. The optimal prediction is given by

$$\begin{aligned} \mathbf{w} &= (\Phi^T \Phi)^{-1} \Phi^T \mathbf{t} \\ \text{and } \Phi &= \begin{pmatrix} \phi_0(x_1) & \dots & \phi_{M-1}(x_1) \\ \phi_0(x_2) & \dots & \phi_{M-1}(x_2) \\ \vdots & \ddots & \vdots \\ \phi_0(x_N) & \dots & \phi_{M-1}(x_N) \end{pmatrix} \end{aligned} \quad (3)$$

The optimal prediction for a new data point x_* is given by:

$$t_* = \mathbf{w}^T \phi(x_*) \quad (4)$$

LSR provides a closed solution to the regression problem. However it does not mirror the significance of differing amounts of data available to derive model parameters. A larger sized dataset with more training examples and stable noise parameters potentially provides more useful information for training a more accurate model than a smaller dataset. LSR is also prone to the presence of outliers because it does not take into account the consistency of

points in a dataset. One or two outliers can sometimes seriously skew LSR results. Another drawback of LSR is its tendency to *over-fit* to a given dataset due to which it often performs poorly on unseen data points. One solution is to include a regularization term λ that controls the relative importance of data-dependent noise. However finding the optimal λ involves techniques such as K-fold cross-validation and the need to maintain a separate validation dataset. These methods can be computationally expensive and wasteful of valuable data. A more elegant solution involves a Bayesian treatment of linear regression. This avoids the over-fitting problem and also permits methods of determining model complexity.

C. Bayesian Linear Regression

Bayesian Linear Regression (BLR) [31] adopts a Bayesian approach to linear regression problem by introducing a prior probability distribution over the model parameters $\mathbf{w}$ in

Equation. 1. Specifically we choose a Gaussian prior over $\mathbf{w}$, $p(\mathbf{w}) = \mathcal{N}(\mathbf{w}; 0, \alpha^{-1} \mathbf{I})$ where α is a hyperparameter.

Given Equation 2, the prior distribution over $\mathbf{w}$, and the properties of Gaussians, we can estimate the posterior distribution of $\mathbf{w}$ given the dataset $\mathcal{D}$ as

$$\begin{aligned} p(\mathbf{w}|\mathbf{t}) &= \mathcal{N}(\mathbf{m}_N, \mathbf{S}_N) \\ \text{with } \mathbf{m}_N &= \beta \mathbf{S}_N \Phi^T \mathbf{t} \\ \text{and } \mathbf{S}_N^{-1} &= \alpha \mathbf{I} + \beta \Phi^T \Phi \end{aligned} \quad (5)$$

The optimal prediction for a new data point is given by the predictive distribution by marginalizing over $\mathbf{w}$ as:

$$p(t_{new}|x_*, \mathbf{t}, \alpha, \beta) = \mathcal{N}(\mathbf{m}_N^T \phi(\mathbf{x}), \sigma_N^2(x_*)) \quad (6)$$

$$\text{and } \sigma_N^2(x_*) = \frac{1}{\beta} + \phi(x_{new})^T \mathbf{S}_N \phi(x_*) \quad (7)$$

In a fully Bayesian approach, we adopt hyperpriors over α and β also and make predictions by marginalizing over $\mathbf{w}$, α and β . However, complete marginalization over all these variables is analytically intractable. We instead adopt an iterative approach by finding the best α and β to maximize the evidence function given this dataset, find the best parameters $\hat{\mathbf{w}}$ to maximize the likelihood given a fixed α and β and repeat until convergence.

The output prediction of BLR as in Equation 6 involves computing a mean $\mathbf{m}_N$ and a variance $\sigma_N^2(\mathbf{x})$. The importance of a variance estimate is that it allows the user to evaluate how “confident” the algorithm is of its prediction. This provides the necessary tools to evaluate the goodness of prediction of an unseen data point. Also, it can be seen from Equation 7 that if an additional point $\mathbf{x}_{N+1}$ were added, the resultant variance

$\sigma_{N+1}^2(\mathbf{x}) < \sigma_N^2(\mathbf{x})$. This tends to the limit $\sigma_{N \rightarrow \infty}^2(\mathbf{x}) = \frac{1}{\beta}$ or the intrinsic noise in the process. Thus BLR reflects the availability of larger quantities of data through a smaller variance and hence an increasingly more confident estimate. The use of a prior hyperparameter α helps guard against over-fitting. It can be seen that α and β are derived purely from the dataset without needing a separate validation dataset. Using BLR helps evaluate whether a probabilistic model helps achieve better prediction accuracy as compared to LSR.

Linear regression models belong to the class of parametric models. Parametric models suffer from a shortcoming in that the form of the basis functions $\{\phi_j(\mathbf{x})\}$ are fixed before the training data set is observed. If the assumption behind the choice of basis functions or the linearity of the model is violated, the model will provide wrong predictions. An alternative to using parametric models is to use nonparametric models where the model structure and complexity are not specified apriori but are instead determined from data.

D. Gaussian Process Regression

Given a set of training points $\{(x_1, t_1), (x_2, t_2), \dots, (x_n, t_n)\}$ such that

$$t_n = f(x_n) + \epsilon, \epsilon \sim \mathcal{N}(0, \beta^{-1} \mathbf{I}) \quad (8)$$

a Gaussian Process Regression model (GPR) [32] estimates a posterior probability distribution over functions $f(x_1), f(x_2), \dots, f(x_N)$ evaluated at points $x_1, x_2, \dots, x_N$ such that any finite subset of the functions is a joint multivariate Gaussian distribution. Consequently, for a given set of points $x = (x_1, x_2, \dots, x_N)^T$, we have a corresponding vector

$\mathcal{F}_x = (f(x_1), f(x_2), \dots, f(x_N))^T$ that belongs to a multivariate Gaussian distribution

$$\mathcal{F}_x \sim \mathcal{N}\{\mu_N(\mathbf{x}), \mathbf{K}_N(\mathbf{x}, \mathbf{x}')\} \quad (9)$$

where $\mu(\mathbf{x})$ is the mean function $\mu(\mathbf{x}) = (\mu(x_1), \mu(x_2), \dots, \mu(x_N))^T$ and $\mathbf{K}$ is the covariance or kernel function. The key idea in GPR is that the covariance between two function values, $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$, depends on the input values, $\mathbf{x}_i$ and $\mathbf{x}_j$ and is specified via the kernel $k(\mathbf{x}_i, \mathbf{x}_j)$. The kernel function returns the covariance between the corresponding $\mathcal{F}_x$ variables $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$. To completely specify a GP, it is enough to specify $\mu(\mathbf{x})$ and $\mathbf{K}(\mathbf{x}, \mathbf{x}')$. By definition, each $f(\mathbf{x}_i)$ is marginally Gaussian, with mean $\mu(\mathbf{x}_i)$ and variance $k(\mathbf{x}_i, \mathbf{x}_i)$.

Typically, for ease of implementation, the mean of the dataset is subtracted each data point so that the mean function is $\mathbf{0}$. To reflect that similar feature vectors with small interpoint Euclidean distance are more likely to correspond to the same output energy consumption measure and to capture the inherent common structure represented by feature vectors due to an underlying periodicity in walking we choose the radial basis function kernel. Also, to capture the fact that we only have access to noisy observations of the function values, it is necessary to add the corresponding covariance function for noisy observations. The complete kernel function can be expressed in element by element fashion as:

$$k(x_i, x_j) = \sigma_f^2 e^{-\frac{1}{2l^2} \|x_i - x_j\|^2} + \sigma_n^2 \delta_{ij} \quad (10)$$

where σ_f^2 is the signal variance, l is a length scale that determines strength of correlation between points, σ_n^2 is the noise variance.

For a new point x_* there exists a corresponding target quantity $f(x_*)$. Since $f(x_*)$ also belongs to the same GP, it can be appended to the original training set to obtain a larger set.

$$\mathcal{F}_{x_{N+1}} \sim \mathcal{N}\{\mu_N(\mathbf{x}), \mathbf{K}_{N+1}(\mathbf{x}, \mathbf{x}')\} \quad (11)$$

$$K_{N+1} = \begin{pmatrix} \mathbf{K}_N & \mathbf{k}_* \\ \mathbf{k}_*^T & k \end{pmatrix} \quad (12)$$

where $\mathbf{k}_*$ has elements $k(x_n, x_*)$ for $n = 1, \dots, N$ and $k(x_i, x_j)$ is defined in Equation 10. Using properties of Gaussians and the definition of GPs, it follows that for a new test point

$p(t_* | x_*, \mathbf{X}, \mathbf{t}) = \mathcal{N}(f(x_*); m(x_*), \sigma^2(x_*))$. Because this joint distribution is Gaussian by definition, we have,

$$m(x_*) = \mathbf{k}_*^T \mathbf{K}_N^{-1} \mathbf{t} \quad (13)$$

$$\sigma^2(x_*) = k - \mathbf{k}_*^T \mathbf{K}_N^{-1} \mathbf{k}_* \quad (14)$$

Thus estimating a target energy from training data amounts to evaluating $\mathbf{K}_N$, $\mathbf{k}$ and c and using values shown in Equations. 13 and 14.

Equations. 13 and 14 summarize the key advantages of GPR. Once again, the use of a probabilistic model to obtain a mean and variance for each prediction allows the user to assess the confidence of each prediction. In contrast to BLR however, GPR is non-parametric, its model complexity increases with larger quantities of data as evident from the increasing size of the kernel matrix. GPR avoids the process of explicitly constructing a suitable feature function space by dealing instead with kernel functions. As the kernel implicitly contains a non-linear transformation, no assumptions about the functional form of the feature space is necessary. This allows us to deal with non-linear maps without having to construct non-linear function spaces. The aim behind using this algorithm was to determine whether using a nonlinear probabilistic map (GPR) offers benefits over a linear probabilistic map (BLR) in terms of increased prediction accuracy.

V. Methods

In this section, we explain the data collection and feature extraction details to obtain training data for subject specific models. Algorithms are run on these extracted data.

A. Hardware Description

Human movement was captured with a modified version of the Sparkfun 6DoF Inertial Measurement Unit (IMU) v4 [33]. Fig. 4 illustrates the hardware used. The v4 provides three axes of acceleration data, three axes of gyroscopic data, and three axes of magnetic data with three sensors: a Freescale MMA7260Q triple-axis accelerometer, set at 1.5 g sensitivity, two InvenSense IDG500 500°/s gyroscopes and the Honeywell HMC1052L and HMC1051Z magnetic sensors. For the purpose of this study only the accelerometer and gyroscopes were used. At the time of this study, the absence of triaxial gyroscopes required that two biaxial gyroscopes be mounted perpendicular to each other and calibrated to function as one gyroscope. Control was through an LPC2138 ARM7 processor. Custom firmware was used on the controller board to stream sensor data continuously. Data were sampled at 100 Hz. The unit used Bluetooth to transmit data to either a nearby PC or mobile phone using the RN41 Bluetooth module set at 115200 bps. Maximum range of the transmitter was found to be 5 m in indoor conditions. The entire system received was powered from a 3.3V rechargeable lithium-polymer battery power supply. The sensor was

encased in a custom-designed harness to be worn on the right-iliac crest. The use of sensors in all three axes allowed the capture of periodicity in all three planes – sagittal, frontal and transverse.

B. Participant Statistics

Seven healthy adults (three men, four women) participated in the study. Height and weight of each participant were recorded using a Healthometer balance beam scale. Fig. 5 describes participant statistics. Informed written consent was obtained from participants and the study was approved by the Institutional Review Board, University of Southern California. Participants walked at 11 predetermined speeds between 2.5 mph and 3.5 mph in intervals of 0.1 mph. Speeds were chosen based on the Compendium of Physical Activities [34]. Rate of oxygen consumption ($\dot{V}O_2$, mL/min) was used as the representation of energy expenditure. This was measured using the MedGraphics Cardio II metabolic system with BreezeSuite v6.1B (Medical Graphics Corporation). The metabolic system outputs data at the frequency of every breath. Before each test, the flow meter was calibrated against a 3.0 L syringe and the system was calibrated against O_2 and CO_2 gases of known concentration. The duration of walking at each speed was 7 minutes with two minutes of changeover time to allow for settling of $\dot{V}O_2$ consumption. For each participant, data was recorded in two sessions with the first session consisting of speeds 2.5 mph, 2.8 mph, 3 mph, 3.3 mph and 3.5 mph and the second session constituting the remaining speeds.

All subjects wore a single inertial sensor above the iliac crest on the right hip. For the sake of comparison, a commercial accelerometer, Actigraph GT1M was also mounted firmly in the same location. The GT1M was set to obtain counts in epochs of 10 seconds. These epochs were later time-synced to features obtained from the IMU. Participants were asked to wear the belt tightly to prevent any slippage. The treadmill used for the experiments was the research quality NordicTrack A2550 PRO. Ground truth for treadmill walking was the displayed treadmill speed. Fig. 11a illustrates the recording procedure [8].

C. Data collection and pre-processing

Each sensor stream from the IMU was passed through a lowpass filter with 3dB cutoff at 20 Hz. This frequency was chosen keeping in mind that everyday activities fall in the frequency range of 0.1-10 Hz [6]. Each stream is divided into ten-second epochs. Within each epoch, the 1024 point normalized FFT was extracted to obtain frequency information. The $\dot{V}O_2$ values from the MedGraphics metabolic system that fell within each epoch were averaged and matched appropriately. These represented the ground truth on which to train data. Thus data for each user consisted of a sequence of epochs, each containing features from the IMU and the average rate of oxygen consumption ($\dot{V}O_2$) for that epoch.

D. Training and Testing Procedure

From each participant's data, a fraction was uniformly sampled and partitioned into training data, the remaining fraction constituting test data. Different models were trained with the same training data but with different feature vectors and candidate algorithms. After each training phase, the algorithm was tested on the test data points to predict $\dot{V}O_2$ values. RMS error was calculated as a measure of accuracy. This was repeated over 10 trials for different randomly sampled data and results averaged. This was repeated for training data percentages from 10% to 90% and constituted a per-subject measure of performance. The results were then averaged over all subjects. To understand the context behind the relative magnitude of the errors, it must be noted that the $\dot{V}O_2$ values were in the range of 400-1000 mL/min.

VI. Results

This section provides a comparative analysis of prediction accuracy of algorithms based on different models. We varied the models along three dimensions. Firstly, we considered the effect of different sensor streams. Our study used two kinds of inertial sensors: triaxial accelerometers and triaxial gyroscopes. Within data from each inertial sensor, we compared the effect of using triaxial information versus uniaxial information for prediction energy consumption. Using the best feature space from each of these comparisons, we compared the utility of accelerometers, gyroscopes and a combined solution using both sensors in terms of prediction accuracy. Secondly, using the best feature space from the study, we compared the relative performance of algorithms measured by prediction accuracy. Finally, we performed an empirical comparison of algorithm run time to provide further insight into algorithm choice based on the trade-off between prediction accuracy and processing capability. The motivation between comparing these models was to understand the issues related to optimal representation of treadmill walking to predict energy consumption given a set of inertial sensors.

It is important to remember the bias-variance trade-offs in choosing the right model [30]. In our study, by choosing a 1024 point Fourier transform of each sensor stream to represent treadmill walking, we choose a high dimensional space of representation. Such a system is prone to over-fitting with a small sized dataset. However, practical considerations related to human participants allow only for the collection of a finite sized dataset. Any approach that aims to be successful at predicting target energy expenditure must prevent over-fitting even with limited data sets. One the other hand, the map from the feature space to the target space is not necessarily linear. To counter a bias towards linear models, one can use nonlinear models to model subtleties. Nonlinear models however are prone to over-fitting when small data sets are used.

A. Comparison across feature spaces

This section deals with choosing the optimal feature space for determining energy expenditure across three algorithms. A comparative study of accelerometer and gyroscope features is shown. Given that gyroscopes have emerged as increasingly promising candidates for tracking human movement, this section also aims at comparing the efficacy of gyroscopes with accelerometers in determining energy expenditure from treadmill walking.

1) Single sensor feature space comparison—Our first task was to identify the optimal feature space identified by that which gives the least average RMS prediction error across all participants when using only one type of sensor is used. Fig. 7 groups the results accordingly. The left and right column of panels contain results when only accelerometers and gyroscopes are used respectively. Each row displays results when one particular algorithm is used in the order - LSR, BLR and GPR. Each panel consists of testing errors obtained when single axes features are used along with a fourth series consisting of triaxial features for comparison.

The first observation was that LSR was sensitive to the amount of training data available regardless of the sensor. For both sensors and LSR, error using single axis streams peaks when 30% of the training data is used. This is not true when triaxial features were used. LSR does not take into account the significance of a data point. Noisy data points have as much weightage as good data points. With lower percentages of training data, the presence of noisy data points biases the predictions made by the algorithm. As more data is available, the biasing effect of noisy data points is reduced which leads to better prediction and lower errors. The use of triaxial features combats this using a richer description with more features. However LSR is also prone to over-fitting. As more training data is available, the

model based on triaxial features begins to over-fit to the dataset. Therefore, with higher percentage of training data, the model based on triaxial feature showed higher errors.

BLR and GPR are less prone to over-fitting at all percentages. BLR achieves this by adopting a Bayesian prior over the weights (Section IV-C). GPR assumes less stringent conditions on data points (only that they be related through a covariance matrix) and hence can handle noisy data points (Section IV-D). Using BLR and GPR, in all feature spaces, increasing the percentage of training data has the effect of reducing prediction errors for that space. For these reasons, in the remainder of this paper, we focus on results obtained from BLR and GPR alone.

In the case of using accelerometer information alone, for both algorithms the errors in increasing order were: Accelerometer all three axes, Accelerometer X axis alone and Accelerometer Z axis or Accelerometer Y axis alone. Using all three axes had the effect of introducing redundancy in trends and thus better prediction accuracies. Thus while individual errors for each axis are higher, combining information introduces results in lower errors. The second best error can be understood by the fact that the dominant acceleration when wearing the sensor in the right hip is in the up-down direction of movement. Movement in this plane represents the best single-axis indicator for predicting energy expenditure. However, triaxial information improved prediction even further. Y axis features represent movement in the forward-backward direction and Z axis features represent movement in the left-right sway direction. The reduction in error when only Y-axis features were used was comparatively less. This could be because of the natural tendency of participants to speed up or down slightly (this constitutes an acceleration or deceleration) while walking to maintain constant position on the treadmill. This would result in a noisy feature vector in this direction. Since the treadmill only allows movement to be forward or backward, the variation for the X and Z axes would be more constrained.

In the case of using gyroscopic information alone, all individual axes showed similar errors. Combining all three axes to obtain triaxial information resulted in higher accuracy. Gyroscopes track rotational rates of the human body rather than accelerations. Hence they were less prone to acceleration in a particular direction. Gyroscopes are also more tolerant to minor changes in position of mounting. Also, gyroscopes capture only dynamic movement free from gravitational bias. All these factors contributed to consistent prediction across all three axes when gyroscopes were used. Increasing the percentage of training data resulted in similar prediction errors for all three individual axes and triaxial features. Thus one method to counter inaccuracies in models is to obtain more training data. Finally, the range of errors obtained when only gyroscopes are used was comparable and in some cases even better than that when only accelerometers were used. This suggests the feasibility in using gyroscopes to track dynamic activities either as a separate sensor or in combination with accelerometers.

2) Comparison between accelerometer and gyroscopic data—Given that the best feature space when using either accelerometer or gyroscopic data only involves using triaxial features, we compared the relative performance of accelerometers and gyroscopes and examined whether combining this information can improve prediction accuracies. Fig. 8 outlines the results when BLR and GPR based models are used. In both cases, using only gyroscope data provided comparable average RMS-errors to using only accelerometer data. Additionally, combining accelerometer and gyroscope information reduces prediction errors. The error reduction obtained from combining sensor information was more pronounced in the case of GPR than LSR. This can be understood from the fact that GPR based models are based on covariances or similarities between data points rather than an explicit dependence on features (Section IV-D). Additional information by way of gyroscopes provides further evidence that a certain data point belongs to a particular class. This enhanced the modeling

capability of GPR. By contrast, BLR is based on obtaining weights for features from training. For the same “kind” of walking the periodicities exhibited would remain the same for both accelerometers and gyroscopes. Thus addition of gyroscopic features would simply amount to a redistribution of the original weights from single sensor data to both sensor streams. It must be noted that in absolute terms, the smallest errors for GPR and BLR were of the same level ($\sim 50 \text{ mL/min}$).

B. Comparison across Algorithms

Given the best feature space, the second study involved determining the algorithm that provided the least Average RMS-error prediction error. Owing to its dependence on data, LSR was ignored and BLR and GPR with triaxial features from both sensors were considered.

1) Algorithm accuracy—Fig. 9 illustrates the results obtained from comparing a nonlinear approach like GPR with a linear approach like BLR. Both GPR and BLR performed better when more training data are used. With increasing training data, GPR performance improved gradually till comparable with BLR. Nonlinear approaches require more data to be able to capture nonlinear subtleties and prevent over-fitting to noise. When data from each subject was considered, GPR showed a lower average RMS prediction error when compared with BLR and LSR when a larger relative percentage of training data was used. This indicates that GPR shows superior performance when a large quantity of data is available. With smaller quantities of data it would be advised to use BLR to prevent over-fitting.

C. Algorithm Run Time

To obtain a more complete picture of algorithmic comparison, we performed a run time analysis of each algorithm with respect to time taken to train on a given dataset. Parametric approaches like linear regression depend on the dimension of the input data space, d and learning is of order $O(d^3)$. Nonparametric approaches depend on the number of data points. In particular, for N data points, GPR requires the inversion of an $N \times N$ matrix which is an $O(N^3)$ operation. Knowing the run time for training is important to understand the trade-offs between prediction accuracy and time of training. This is particularly important if these algorithms are to be implemented in resource-constrained platforms such as mobile phones or portable PCs. In our study, there were three classes of data types: Single sensor (either accelerometer or gyroscope) with only one axis in use, single sensor with all three axes in use and both sensors all three axes in use. Each of these cases multiplies the feature space used by 3. In addition, three algorithms: LSR, BLR and GPR were used. Fig. 10 illustrates our results for one participant. Similar trends exist for all participants.

For this study, the time taken to train a dataset with different percentages of training data for one participant was recorded in the case of one feature space and one algorithm. Prediction accuracies were also measured. A scatter plot was made with prediction accuracies on the X-axis and algorithm run time on the Y-axis (Log-scale, base 10) with all training percentages represented as one class. This was repeated for different combinations of feature vectors and algorithms. In all plots, feature spaces are coded by color (Blue: Single sensor, single axis; Red: Single sensor, all axes; Green: Both sensors, all axes) and algorithms are coded by symbol (LSR: Empty square; BLR: Filled circle; GPR: Empty Star). For clarity, plots are shown in two views.

Fig. 10a shows a comparative analysis of run time versus accuracy for BLR and LSR. In the case of LSR, addition of extra features showed no benefits in terms of accuracy but increases run time. This is because LSR is prone to over-fitting with high dimensional

spaces and is easily affected by noise. Addition of features improved BLR prediction accuracy at the expense of higher run time. However in our study, the absolute run time for training was still on the order of a few seconds and less than 30 seconds in all cases for all participants. The consistency of results and lower error rates along with reasonable training run times justifies the selection of BLR over LSR with any combination of feature vectors. Using both sensors offers limited advantage in terms of prediction accuracy but requires larger run times. Therefore, it would be advisable in resource-constrained systems to choose a model that only uses either sensor for training if accuracy is not an issue.

Fig. 10b shows a comparative analysis of run time versus accuracy for BLR and GPR. Both BLR and GPR are probabilistic approaches and hence show consistently better results with increasing dataset size. Nonlinear modeling with GPR showed lower or comparable errors as BLR, particularly when more training data were used. However, the run time was at least two or three orders of magnitude higher. To provide intuition behind such run time, we observe that a $\dot{V}O_2$ estimation error of 35 mL/min (best accuracy possible in our study with BLR for this participant) corresponds roughly to a percentage error of 5% or lower. To go from 35 mL/min to 30 mL/min (best accuracy possible with GPR for this participants) corresponds to obtaining a percentage error of 4% or lower. Given a dataset, to be able to obtain a higher accuracy requires increasingly larger computing power to accommodate more sophisticated models. In resource-constrained systems, this incremental increase in accuracy might not be justified. Therefore if computing power is an issue, it would be advisable to use linear models like BLR over nonlinear models.

D. An example prediction

Fig. 11a illustrates an example output for energy prediction for a single participant using the feature space and algorithm combination that provides the lowest prediction error. The model uses triaxial features from both accelerometer and gyroscope axes and Gaussian process regression with 80% of the data used for training. The predicted values (shown in blue) closely match the ground truth (shown in red). Fig. 11b shows the same data as a scatter plot with ground truth displayed on the X-axis and GPR predicted values on the Y-axis. The two time series showed a high positive correlation of 0.92 across users. In order to compare it with existing standards, we also mounted a commercial accelerometer, the Actigraph GT1M on the same location as IMU. The GT1M provides counts as a measure of energy expenditure. The counts output by the GT1M only showed a correlation of 0.85 for the X-axis and .87 for the Y-axis for the same data points. This showed the advantage of the combination of raw data, extracting features based on some property of treadmill walking and probabilistic modeling of the regression problem.

VII. Conclusion and Future Work

Walking is one of the easiest and commonly available activities that can be used to maintain an active lifestyle. Being able to accurately characterize intensity of walking represents an important step in public health delivery because it allows development of appropriate intervention measures based on objective measures. Existing techniques using commercial accelerometers rely on “counts” to represent activity information. This suffers from two drawback - imprecision because of proprietary counts and incompleteness because human movement cannot be completely represented without gyroscopes. Being able to accurately represent human activity and derive physiological measures such as energy expenditure requires that one extracts meaningful and complete properties of human movement from raw data and develop relevant maps based on these properties.

Our approach

In this paper, we described an experimental study in estimating energy expenditure from treadmill walking using a single hip-mounted triaxial inertial sensor comprised of a triaxial accelerometer and a triaxial gyroscope. Our approach involved representing the cyclic nature of walking using Fourier transforms of triaxial accelerometer and gyroscopic sensor streams and showing the relationship between Fourier domain features and energy expended. We then showed how one can develop a probabilistic map from these representations to energy expenditure (as measured by $V O_2$ consumption, mL/min). We described three regression techniques: Least Squares Regression (LSR), Bayesian Linear Regression (BLR) and Gaussian Process Regression (GPR). Least Squares regression is the simplest and most direct form of linear regression techniques. Bayesian Linear Regression represents a probabilistic version of ordinary Least Squares Regression obtained by defining a distribution over LSR weights. Gaussian Process Regression is a non-linear regression technique that defines a joint probability distribution over functional mappings. We performed two studies. The first was a comparative analysis of different feature combinations of sensor streams from triaxial accelerometers and gyroscopes to find the most accurate representative space to predict energy expenditure (measured by RMS prediction accuracy) for treadmill walking. The second study was an analysis of the best algorithmic approach among linear and nonlinear methods as measured by RMS prediction accuracy and run time.

Results

Our results detailed prediction accuracies (measured by Average RMS prediction errors) obtained using different sensor streams and algorithmic approaches. LSR results depended heavily on the number of points used for training. This was because LSR is prone to over-fitting and the presence of outliers. BLR and GPR showed consistently reducing errors with increasing training data size. When BLR and GPR were considered, using accelerometer and gyroscope data simultaneously improved prediction accuracy as opposed to using just one uniaxial features. Among accelerometer features using X-axis acceleration showed the next lowest prediction error. This was because the X-axis was aligned with the direction of forward movement. All three gyroscopic axes showed comparable errors. Using gyroscope features in Y and Z axes shows lower errors than corresponding acceleration Y and Z axes. This is due to gyroscopes only capturing dynamic rotational movement free from gravitational bias. Gyroscopes were capable of providing equivalent if not better results for energy prediction from treadmill walking. Additionally, combining accelerometer and gyroscope information reduced prediction errors. Both GPR and BLR performed better when more training data were used. With increasing training data, GPR performance improved till comparable with BLR highlighting the need for more data for nonlinear approaches. However, GPR training run time was at least two or three orders of magnitude higher. Therefore if computing power is an issue, it would be advisable to use linear models like BLR over nonlinear models with a tradeoff in accuracy.

We aim to expand our work in a number of directions. We are currently working on developing generalized models that are applicable across a wide range of physiological parameters such as height, weight, BMI, gender and age. This involves undertaking a larger study and collecting movement information for treadmill walking. We aim to develop a generalized linear model similar to Bayesian Linear Regression with informative priors. Secondly, we also plan to undertake similar analyses for other kinds of cyclic activities. In doing so, we aim to explore whether our methods can be applied in the more general framework of energy expenditure for repetitive activities.

Acknowledgments

The authors would like to thank David Erceg of the Division of Biokinesiology and Physical Therapy, USC, for his invaluable support and guidance.

This work was supported in part by Qualcomm, Nokia, NSF (CCR-0120778) as part of the Center for Embedded Network Sensing (CENS), and the USC Comprehensive NCMHD Research Center of Excellence (P60 MD 002254). Support for H. Vathsangam was provided by the USC Annenberg Doctoral fellowship program.

References

- [1]. World Health Organization - Physical Activity. January. 2010
<http://www.who.int/dietphysicalactivity/pa/en/index.html>
- [2]. Stewart ST, Cutler DM, Rosen AB. Forecasting the effects of obesity and smoking on u.s. life expectancy. *N Engl J Med.* 2009; 361(23):2252–2260. [PubMed: 19955525]
- [3]. Warburton DE, Whitney C, Bredin SS. Health Benefits of Physical Activity. *CMAJ.* 2006; 174
- [4]. Aminian K, Najafi B. Capturing human motion using body-fixed sensors: outdoor measurement and clinical applications. *Computer Animation and Virtual Worlds.* 2004; 15(2):79–94.
- [5]. Meijer G, Westerterp K, Verhoeven F, Koper H, ten Hoor F. Methods to assess physical activity with special reference to motion sensors and accelerometers. *Biomedical Engineering, IEEE Transactions on.* March.1991 38:221–229.
- [6]. Bao, L.; Intille, SS. Activity recognition from user-annotated acceleration data. April. 2004
- [7]. Dunn AL, Marcus BH, Kampert JB, Garcia ME, Kohl HW III, Blair SN. Comparison of Lifestyle and Structured Interventions to Increase Physical Activity and Cardiorespiratory Fitness: A Randomized Trial. *JAMA.* 1999; 281(4):327–334. [PubMed: 9929085]
- [8]. Vathsangam, H.; Emken, A.; Spruijt-Metz, D.; Schroeder, TE.; Sukhatme, GS. Energy Estimation of Treadmill Walking using Onbody Accelerometers and Gyroscopes; 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2010;
- [9]. Bouten C, Westerterp K, Verduin M, Janssen J. Assessment of energy expenditure for physical activity using a triaxial accelerometer. *Med Sci Sports Exercise.* 1994; 26(12):1516–1523.
- [10]. Mathie MJ, Coster AC, Lovell NH, Celler BG. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiological measurement.* April.2004 25
- [11]. Crouter SE, Clowers KG, Bassett DR. A novel method for using accelerometer data to predict energy expenditure. *J Appl Physiol.* Apr.2006 100:1324–1331. [PubMed: 16322367]
- [12]. Freedson PS, Melanson E, Sirard J. Calibration of the computer science and applications, inc. accelerometer. *Medicine and science in sports and exercise.* May.1998 30:777–781. [PubMed: 9588623]
- [13]. Klippel N, Heil D. Validation of energy expenditure prediction algorithms in adults using the actical electronic activity monitor. *Med Sci Sports Exerc.* 2003; 35
- [14]. Reilly JJ, Kelly LA, Montgomery C, Jackson DM, Slater C, Grant S, Paton JY. Validation of actigraph accelerometer estimates of total energy expenditure in young children. *Int J Pediatr Obes.* 2006; 1(3):161–167. [PubMed: 17899634]
- [15]. Plasqui G, Westerterp KR. Physical activity assessment with accelerometers: An evaluation against doubly labeled water. *Obesity.* Oct..2007 15:2371–2379. [PubMed: 17925461]
- [16]. Troiano RP. Translating accelerometer counts into energy expenditure: advancing the quest. *J Appl Physiol.* 2006; 100:1107–1108. [PubMed: 16540708]
- [17]. Heil DP. Predicting activity energy expenditure using the actical? ® activity monitor. *Research Quarterly for Exercise and Sport.* March.2006 77:64. [PubMed: 16646354]
- [18]. Godfrey A, Conway R, Meagher D, Laighin G. Direct measurement of human movement by accelerometry. *Medical Engg. & Phy.* 2008; 30:1364–1386.
- [19]. Tapia, M. PhD thesis. Massachusetts Institute of Technology. Dept. of Architecture. Program in Media Arts and Sciences; 2008. Using machine learning for real-time activity recognition and estimation of energy expenditure.

- [20]. Rothney M, Neumann M, Beziat A, Chen K. An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *J Appl Physiol.* October.2007 103:1419–27. [PubMed: 17641221]
- [21]. Albinali, F.; Intille, S.; Haskell, W.; Rosenberger, M. Using wearable activity type detection to improve physical activity energy expenditure estimation; 12th conference on Ubiquitous Computing; 2010;
- [22]. Rowlands AV. Accelerometer assessment of physical activity in children: an update. *Pediatr Exerc Sci.* Aug.2007 19:252–266. [PubMed: 18019585]
- [23]. Cheng, J-C.; Moura, JMF. Tracking human walking in dynamic scenes; ICIP '97: Proceedings of the 1997 International Conference on Image Processing (ICIP '97) 3-Volume Set-Volume 1; Washington, DC, USA. 1997; p. 137IEEE Computer Society
- [24]. Zappa, B.; Legnani, G.; van den Bogert, AJ.; Adamini, R. On the number and placement of accelerometers for angular velocity and acceleration determination. 2001.
- [25]. Single-chip triaxial gyro from invensense. 2009.
<http://www.sensormag.com/product/single-chip-triaxial-gyro-invensense>
- [26]. Chang C, Ansari R, Khokhar A. Efficient tracking of cyclic human motion by component motion. 2004; (12):941–944.
- [27]. ACSM's Guidelines for Exercise Testing and Prescription. American College of Sports Medicine; 2010.
- [28]. Bunc, V.; Dlouha, R. Energy cost of treadmill walking. *Anglais:* 1997.
- [29]. Pearce M, Cunningham D, Donner A, Rechnitzer P, Fullerton G, Howard J. Energy cost of treadmill and floor walking at self-selected paces. *European Journal of Applied Physiology and Occupational Physiology.* 1983; 52:115–119. 10.1007/BF00429037. [PubMed: 6686120]
- [30]. Bishop, CM. Pattern Recognition and Machine Learning (Information Science and Statistics. Springer-Verlag; 2006.
- [31]. Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. Bayesian Data Analysis. Second Edition. Chapman and Hall/CRC; July. 2003 2 ed.
- [32]. Williams, CKI. Prediction With Gaussian Processes: From Linear Regression To Linear Prediction And Beyond. 1997. p. 599-621.
- [33]. Sparkfun. 2009. "6 Dof v4 Datasheet -
<http://www.sparkfun.com/datasheets/Sensors/IMU/6DOF-v4-Rev1.pdf>,"
- [34]. Ainsworth B, Haskell W, Whitt M, Irwin M, Swartz A, O'Brien SSW, K. Schmitz DB Jr, Emplaincourt P, Leon A. Compendium of physical activities: An update of activity codes and met intensities. *Med Sci Sports Exerc.* 2000; 32:S498–S516. D. J. Jr. [PubMed: 10993420]

Figure 1.

An example of periodic signals obtained from the iliac crest of the right hip while walking at a constant speed of 2.5 mph. Steady state walk is cyclic in nature and this periodicity can be captured with inertial sensors.

Figure 2.

The Fourier transforms of the accelerometer stream on the X-axis for steady state walking at 2.5 mph and 3.5 mph. Definite peaks characteristic of walking speed suggest energy estimation by monitoring the frequency spectrum

Figure 3.

A color map illustrating the variation of Fourier transform feature vectors as a function of speed. The map shows a set of dominant frequencies that occur at regular intervals. These dominant frequencies shift to the right with increasing speed. This is indicative of higher frequency of step counts during faster walk. The relationship between energy expenditure and walking speed has often been modeled as one where walking at higher speeds requires higher energy expenditure

Figure 4.

Illustration of hardware used to capture treadmill walking information. Acceleration information was collected with a Freescale MMA7260Q triple-axis accelerometer. Rotational rates were collected with 2 Invensense IDG500 $500^{\circ}/s$ gyroscopes mounted perpendicular to each other. Data was transmitted via Bluetooth with the Roving Networks RN41 Bluetooth module operating at 100 Hz on a 115200 bps link. The sensor hardware was modified to be worn with a custom designed harness on the right iliac crest. Original image source: www.sparkfun.com

Figure 5.

Statistics of participant population. Seven participants: 3 males (Blue squares) and 4 females (Green circles) were involved in the study. Average age = 29 ± 6 years. Average Height = 1.67 ± 0.10 m. Average Weight = 66 ± 17 kg. Average BMI = 24 ± 8 .

Figure 6. An example recording procedure for a single participant. The yellow box indicates sensor mounting. The red box indicates $V O_2$ recording via the mask leading to the metabolic cart.

Figure 7.

Illustration of variation of prediction accuracy (measured by Average RMS prediction error across all participants) with different combinations of feature vectors. Results are grouped row-wise by algorithm and column wise by sensor stream. LSR results depended heavily on the number of points used. This is because LSR was prone to over-fitting and presence of outliers. Over-fitting is more common with high-dimensional spaces and low number of data points. BLR and GPR showed consistently reducing errors with increase in training data size. In the case of BLR and GPR, use of all 3 axes as features improved prediction accuracy as opposed to using just one sensor axis. Among accelerometer features using X-axis acceleration alone showed the next lowest prediction error. This is because the X-axis was aligned with the direction of forward movement. All three gyroscopic axes showed comparable errors. Using gyroscope features in Y and Z axes is showed lower errors than corresponding acceleration Y and Z axes. This was due to gyroscopes only capturing dynamic movement free from gravitational bias. Gyroscopes were capable of providing equivalent if not better results for energy prediction from treadmill walking.

(a) Accelerometer and gyroscopic information show similar prediction accuracies. Combining accelerometer and gyroscopic information shows lower prediction errors.

(b) Combining accelerometer and gyroscopic information shows lower prediction errors. The decrease in error is higher than in the case of BLR because GPR predicts information based on covariances between data points and hence incorporates relationships between similar data points as well. Additionally, it is a nonlinear method.

Figure 8.

Illustration of the effect of combining triaxial accelerometer and gyroscopic information (measured by Average RMS prediction error across all participants) in the case of BLR and GPR. Accelerometer and gyroscope provide similar results when used separately.

Figure 9.

Illustration of relative algorithmic performance when triaxial information from all sensors is used (measured by Average RMS prediction error across all participants). With increasing number of data points GPR begins to perform comparably with BLR. The accuracy of nonlinear techniques is predicated by the availability of large amounts of data. Larger data allow the recognition of subtleties in the data by understanding the distinction between the non-linearities in feature space and noise. With larger data sets, nonlinear techniques are justified, whereas with small data sets, it is recommended to use linear techniques to avoid over-fitting.

Figure 10.

An illustration of the relationship between accuracy and run time for LSR, BLR and GPR for a single participant. Scatter points of each class represent different training percentages of the same class and feature space. The best algorithm has to be as close to the origin as possible (lowest error and lowest run time)

(a) An example prediction for a randomly selected test dataset from a single subject. Sample number is on the X-axis. The ground truth is displayed in red and output from GPR in blue with associated prediction variance in gray. Predicted values closely match the ground truth as obtained from the Medgraphics metabolic cart. The advantage of using probabilistic measures like GPR is the usage of a variance measure to assess the confidence of a prediction. Similar plots can be observed for other subjects and other combinations of features and algorithms.

(b) An example scatter plot showing GPR prediction versus ground truth obtained for a single participant. Predicted points showed a high positive correlation of 0.92 across subjects indicating the prediction accuracy versus using just counts. By contrast, a commercial accelerometer the Actigraph GT1M only showed a correlation of 0.85 for the X-axis and .87 for the Y-axis for the same data points.

Figure 11.

An example prediction using the algorithm and feature combination that provided the least error across subjects. The prediction is based on a learned model using all sensor streams and all axes with Gaussian Process Regression (GPR) as the training algorithm.