Optimal Time-Resource Allocation for Energy-Efficient Physical Activity Detection

Gautam Thatte*, Ming Li, Sangwon Lee, Adar Emken, Murali Annavaram,

Shrikanth Narayanan, Donna Spruijt-Metz, and Urbashi Mitra

Abstract

The optimal allocation of samples for physical activity detection in a wireless body area network for health-monitoring is considered. The number of biometric samples collected at the mobile device fusion center, from both device-internal and external Bluetooth heterogeneous sensors, is optimized to minimize the transmission power for a fixed number of samples, and to meet a performance requirement defined using the probability of misclassification between multiple hypotheses. A filter-based feature selection method determines an optimal feature set for classification, and a correlated Gaussian model is considered. Using experimental data from overweight adolescent subjects, it is found that allocating a greater proportion of samples to sensors which better discriminate between certain activity-levels can result either a lower probability of error or energy-savings ranging from 18% to 22%, in comparison to equal allocation of samples. The current activity of the subjects and the performance requirements do not significantly affect the optimal allocation, but employing personalized models results in improved energy-efficiency. As the number of samples is an integer, an exhaustive search to determine the optimal allocation is typical, but computationally expensive. To this end, an alternate, continuous-valued vector optimization is derived which yields approximately optimal allocations and can be implemented on the mobile fusion center due to its significantly lower complexity.

EDICS Category: BIO-MAID

G. Thatte, M. Li, S. Lee, U. Mitra, S. Narayanan and M. Annavaram are with the Ming Hseih Department of Electrical Engineering, and A. Emken and D. Spruijt-Metz are with the Keck School of Medicine, University of Southern California, Los Angeles, CA. e-mail: {thatte,mingli,sangwonl,emken,ubli,annavara,dmetz}@usc.edu, shri@sipi.usc.edu. 3740 McClintock Ave., Room 522, Los Angeles, CA 90089-2565. Phone: 213-740-4684, Fax: 213-740-4449

I. INTRODUCTION

Wearable health monitoring systems coupled with wireless communications are the bedrock of an emerging class of sensor networks: wireless body area networks (WBANs). Such networks have myriad applications, including diet monitoring [38], detection of activity or posture [4, 37], and health crisis support [15]. This paper focuses on the KNOWME network, which is targeted to applications in pediatric obesity, a developing health crisis both within the US and internationally. To understand, treat, and prevent childhood obesity, it is necessary to develop a multimodal system to track an individual's level of stress, physical activity, and blood glucose, as well as other vital signs, simultaneously. Such data must also be anchorable to context, such as time of day and geographical location. The KNOWME network is a first step in the development of a system that could achieve these targets.

The KNOWME system is an end-to-end mobile health platform that interfaces wireless sensors with a Nokia N95 mobile phone (the fusion center) via Bluetooth to precisely monitor heart rate using electrocardiograph (ECG) sensors, blood oxygen levels using a pulse oximeter, and motion using both the mobile phones built-in accelerometer and an external accelerometer. Our current implementation also collects information from other phone-internal sensors including Global Positioning System (GPS) measurements and audio and video tags, which we plan to incorporate into our state detection algorithms in the future.

A crucial component of the KNOWME network is the unified design and evaluation of multimodal sensing and interpretation, which allows for automatic recognition, prediction, and reasoning regarding physical activity and other sensed emotional or behavioral states. This accomplishes the current goals of observational research in obesity and metabolic health regarding physical activity and energy expenditure (traditionally carried out through careful expert human data coding), as well as enabling new methods of analysis previously unavailable, such as incorporating data on the user's emotional state. Our platform develops real-time measurement of physical activity and providing immediate feedback, which can be used in personalized interventions that are tailored to the individual needs of the subject [7].

The Bluetooth standard for data communication, employed by the KNOWME network, uses a "serve as available" protocol in which all samples taken by each sensor are collected by the fusion center. Though this is beneficial from the standpoint of signal processing and activity-level detection, it requires undesirably high energy consumption: with a fully charged battery, the Nokia N95 cellphone can perform

December 26, 2010

over ten hours of telephone conversation, but if the GPS receiver is activated, the battery is drained in under six hours [50]. A similar decrease in battery life occurs if Bluetooth is left on continuously; we quantify the energy consumption of the mobile device in Section III. One of the aims of this paper is to devise a scheme that reduces the Bluetooth communication, thus resulting in energy savings.

Our earlier work [48] suggests that data provided by certain types of sensors are more informative in distinguishing between certain activities than other types. For example, the ECG sensor is a better discriminator when the subject is lying down, or in other activities that require low levels of energy expenditure, while data from the accelerometer is more pertinent to distinguishing between non-sedentary activities, or activities that demand higher energy expenditure. In this paper, activity-detection is considered as a multiple hypothesis testing problem, and the performance of the system is measured via an upper bound on the probability of misclassification. In particular, we consider the optimal allocation of samples amongst heterogeneous sensors, some of which communicate their measurements using Bluetooth to the Nokia N95 fusion center, and specifically focus on two optimization problems that are detailed in Section V[1]. In addition to the external Bluetooth sensors, our algorithms also considers device-internal sensors since their measurements are available at a fraction of the transmission cost of Bluetooth samples. The transmission power cost, incorporating both device-internal and external Bluetooth samples, is explicitly considered in the second of our two optimizations. In particular, the second optimization minimizes the total transmission power cost, which is motivated by the increased power consumption at the Nokia N95 due to continuous, multiple Bluetooth connections to the heterogeneous sensors. Furthermore, as several features can be extracted from individual biometric samples, but employing all the features can be computationally expensive, we implement filter-based feature selection using the symmetrical uncertainty metric (see Appendix A).

The goal of this work is to derive an algorithm for energy-efficient physical activity detection and its low-complexity implementation. Our main contributions are summarized as:

1) Investigating the trade-off between system performance, defined as a function of the probability of misclassification, and energy-consumption due to the transmission of measurements from the sensors to the fusion center.

---

[1] One of the two optimization problems has been previously considered in our earlier work [48], which presented a simpler version of the models developed herein. We note that the earlier work used a single exemplary feature from each sensor; in contrast, this work considers multiple features and employs feature selection to choose an appropriate set of features.

2) Developing a low-complexity implementation that circumvents a typical exhaustive search, and allows us to implement optimal sampling on the mobile device.

3) Testing our algorithms using real data collected using overweight adolescent test subjects, and using personalized training and model parameters for our detection and optimal sampling algorithms.

The remainder of this paper is organized as follows. Prior relevant work on activity-level detection and energy-efficient algorithms in WBANs, and their relationship to our work is presented in Section II. The system architecture of the KNOWME network and the energy consumption due to data collection in the mobile device, which motivates our optimal allocation algorithm, are presented in Section III. Section IV overviews the feature extraction and feature selection methods (detailed in Appendix A), and then outlines the (first-order autoregressive) AR(1)-correlated multivariate Gaussian models. The optimal sample allocation problem, and the low-complexity optimization, is derived in Section V. Section VI presents a performance analysis that is based on data collected using overweight adolescent test subjects. Conclusions and avenues for future work are discussed in Section VII.

II. RELATED WORK

In this section, we review the prior art in activity-level detection and energy-efficient algorithms in WBANs. The research considered can be broadly classified into accelerometry-based systems and multimodal systems, and systems that explicitly implement energy-efficient algorithms.

Accelerometry-based systems: Several projects investigating activity-level detection centers on tri-axial accelerometer data alone (*e.g.* [21, 25, 35, 37, 43]) with some systems employing several accelerometer packages and/or gyroscopes. For example, a multi-accelerometer system that discriminates between various postures by initially differentiating between high- and low-level activities, and further classifies postures using a Hidden Markov Model (HMM) framework is proposed in [37]. Accelerometer-based systems for activity-detection have also been developed for specific applications such as manual wheelchair use [18] and for stroke patients [41]. In contrast to these works, the KNOWME network employs heterogeneous sensors for multimodal sensing in an activity-detection framework.

Multimodal systems: Augmenting accelerometers with other sensors has yielded multi-sensor systems that have been implemented and deployed for activity-level detection. A context-aware wearable system that uses a SenseWear armband with multi-modal sensors to identify various daily activities using the k-Means clustering algorithm is developed in [26]. A similar set of activities are discussed in [3], using

December 26, 2010

a combination of an ear worn sensor and ambient sensors placed around the environment, and employing a two-stage Bayesian classifier with multivariate Gaussian models. In yet another work, various daily activities, as well as certain sporting activities, are detected using accelerometers and GPS measurements via decision tree and neural network classifiers [13].

In addition to activity-detection systems, several multi-sensor systems have been developed for biometric identification [49], context-aware sensing and specific health-monitoring applications. For example, some systems are tailored for emergency response and triage situations [15], while others have developed a lightweight embedded system that is primarily used for patient monitoring [9]. The system developed in [23] is used in the assistance of physical rehabilitation, and the UbiFit system [8] is designed to promote physical activity and an active lifestyle. In these multi-sensor works, emphasis is on higher layer communication network processing and hardware design. However, our work explicitly focuses on developing the statistical signal processing techniques required for activity-level detection. Furthermore, the energy-efficiency of the system is explicitly considered in the KNOWME system, in contrast to the aforementioned works.

Energy-efficient systems: The notion of designing energy-saving strategies, well-studied and implemented in the context of traditional sensor and mobile networks [6, 42], has also been incorporated into WBANs for activity-level detection. For example, the goal of the work in [4] is to determine a sampling scheme (with respect to frequency of sampling and sleeping/waking cycles) for multiple sensors to minimize power consumption. A similar scheme that minimizes energy consumption is based on redistributing un-used time over tasks as evenly as possible [34]. Energy-efficient policies for sensor networks also include hierarchical schemes as described in [11, 17]. In these schemes, a set of low-power sensors remain active continuously, and once an event is detected, other high-power and more sensitive sensors wake up and sample the environment. Our work offers a new approach in that the energy-efficiency of the system is a result of optimized resource allocation; *i.e.* measurements are distributed among sensors according to the sensor that is most informative in a given situation, rather than being equally distributed at all times.

III. SYSTEM OVERVIEW

In this section, we first outline the KNOWME network architecture and multi-hypothesis testing problem, which serves as the framework for the optimal sampling algorithm. We then give an overview

of the overall system design and describe the interactions of the component subsystems, and finally quantify the energy consumption when using the KNOWME platform, which motivates the necessity of an energy-efficient mechanism for physical activity detection.

A. System Architecture

A WBAN node in the KNOWME network consists of the Nokia N95 fusion center, which has an in-built tri-axial accelerometer (denoted NOK) that samples at 30 Hz, and Bluetooth-enabled oximeter (OXI), electrocardiograph (ECG) and tri-axial accelerometer (ACC). The ECG and ACC Bluetooth sensors sample at 300 Hz and 75 Hz, respectively; sensor data is transmitted via Bluetooth to the Nokia N95. The mobile device can also collect GPS measurements and audio/video tags, but these are not actively being used in our applications. Similarly, samples from the OXI are not currently being used since we have found they are significantly less effective, compared to the ACC and ECG, for the detection of physical activities and optimal allocation of samples.

Features extracted and selected from the three sensors (NOK, ACC, ECG), summarized in Section IV-A and detailed in Appendix A, are used to determine the optimal allocation of samples, wherein the goal of the optimization is to minimize the transmission cost while maintaining a predetermined performance level, which is defined via a probability of misclassification (see Section V). In particular, we consider an M-ary hypothesis testing problem wherein each distinct activity maps to a hypothesis. We consider the following eight physical activities: lying down, sitting, sitting and fidgeting, standing, standing and fidgeting, slow walking, brisk walking and running. Our methods are easily extensible to more states and more sensors and features.

B. System Design

The current implementation of the KNOWME system consists of the hardware components, outlined in the previous section, and two algorithmic components: classification algorithms and the optimal sampling algorithm. The former component has been presented in another work [32], and the current work focuses on the design of an optimal sampling algorithm. The data collection procedure, described in Section VI-A, is comprised of 3 to 4 sessions, one of which is used as a training session to develop the higher-dimensional and Gaussian models for the classification and optimal sampling algorithms, respectively.

The classification algorithms currently use all extracted temporal and cepstral features in the testing phase, and the models developed using the training data, to classify the activities using support vector machine (SVM) and Gaussian mixture model (GMM) classifiers [39]. On the other hand, the optimal sampling algorithm employs filter-based feature selection (see Section IV-A and Appendix A), choosing features with strong feature-class correlations and weak feature-feature correlations (see Section VI-B2), to enable further energy-savings while determining the optimal allocation of samples and maintaining the required detection performance. Furthermore, a low-complexity optimization is derived (see Section V-A) that replaces the exhaustive search with a vector optimization, enabling real-time operability of the optimal sampling algorithm.

C. Energy Consumption due to Data Collection

In the KNOWME system, the external sensors simply transmit data to the mobile phone fusion center via the Bluetooth protocol; the Nokia N95 performs all the coordination, processing and computation tasks. Furthermore, the energy of the Nokia N95 is consumed in sensing and collecting samples for the device-internal sensors. Since the internal sensing and Bluetooth transmission energy costs are unequal, there exists a need to examine and optimize the energy consumption in order to extend the battery-life of the mobile device.

Since we consider the optimal allocation of the decoded and processed samples in the following sections, the energy consumption for each of the sensors has been experimentally determined. The device-internal NOK sensor consumes 0.063 W, and the Bluetooth transmission of samples from the external ECG and ACC sensors requires 0.108 W and 0.084 W, respectively. We note that the transmission powers for the two sensors are different since the energy consumed for Bluetooth transmission depends on the data rate; we have assumed that the header overhead for both the ECG and ACC sensors is identical. The varying energy costs for each of the sensors directly affects the result of the optimization, which minimizes this transmission cost while maintaining a performance requirement.

IV. SYSTEM MODEL

In this section, we first provide an overview of the feature extraction and filter-based selection (detailed in Appendix A). We then present the autoregressive (AR) Gaussian model for the biometric features.

December 26, 2010

- mean absolute deviation
- $(20, 40, 60, 80)^{th}$ percentile
- cross correlation with other axes
- inter-quartile range
- mean of maxima   • energy
- standard deviation   • kurtosis
- mean crossing rate   • median
- mean of minima   • mean
- root mean square   • skewness

Box 1: Accelerometer features (averaged across three axes).

A. Feature Extraction and Selection

A feature is a characteristic measurement, extracted from the input data, that represents important patterns of desired phenomena (different physical activities in our case) with reduced dimension [10]. Utilizing the complementary characteristics of different features can offer substantial improvement in the recognition accuracy, while reducing the computational complexity; this is termed *feature selection* and is detailed for our framework in Appendix A. In the case of the ECG, the features extracted are the median and standard deviation of the inter-peak period of the Electrocardiogram (ECG) waveform, a noise measure that accounts for variations in the recorded ECG signal, and coefficients of a Hermite polynomial expansion (HPE) [33], which reconstructs the ECG signal.

We refer the readers to [32] for further details regarding feature extraction from the ECG signal.

For the accelerometer, the features extracted are statistical measures (averaged over the three axes) of overlapping windows of the accelerometer signal, for both the external and phone-internal accelerometers. The average values of the statistical measures over the three axes are employed to reduce dependency on the orientation of the devices. Using the averaged values instead of the individual values affects the feature selection process (reflecting the loss of information due to averaging), but there is no appreciable change in the optimal allocation of samples when using the average of the accelerometer axes. A 6.72-second window (corresponding to 504 samples given the 75 Hz sampling rate for the ACC) with a 50% overlap is used [32], and the statistical measures considered are listed in Box 1.

We employ sequential forward[2] selection (SFS) [51], which is a correlation-based feature selection

[2]The choice of a forward feature selection framework, as compared to a backward, floating or forward-backward selection process, is merely a design choice, and does not impact the optimal allocation that is derived using the selected features.

method, and is detailed in Appendix A. As desrcibed in Section III-B, feature selection is employed to achieve further energy-savings by reducing the computational burden of processing all (and possible redundant) extracted features. The SFS algorithm is a "greedy" algorithm that starts with an empty set, and continues to include features until a metric associated with the current subset does not increase. The performance of the algorithm depends on the correlation metric employed, and we consider the *symmetrical uncertainty* (SU) metric [53]. Appendix A details the feature selection algorithm and presents the results for the extracted ACC and ECG features for the SU metric.

B. AR(1)-Correlated Gaussian Model

We model the extracted features using an AR(1)-correlated multivariate Gaussian model, and explicitly examine the case wherein multiple features from a single sensor are considered. We assume that the individual features employed are uncorrelated. This assumption is consistent with the operation of the various feature selection methods we have employed wherein features that are highly correlated with a particular activity are chosen, but have lower correlation with the other selected features. A consequence of our feature selection is that features from distinct sensors are uncorrelated as are features drawn from the same sensor. The lack of correlation between sensors and between features enables the derivation of a low-complexity implementation of the optimal sampling algorithm. However, assuming independence in the presence of correlations between different sensors and different features is a sub-optimal approach, and results in a marginal loss of performance. But this limitation of our current work enables the development of the low-complexity implementation that replaces an exhaustive search with a vector optimization, which results in being able to implement the optimal sampling algorithm in real time on the mobile device (see Section V-A for details). Our assumption is borne out by our numerical results (see Section VI-B). While individual features are uncorrelated, there is temporal correlation for a single feature.

Thus, we consider the following signal model for the decoded and processed samples, corresponding to the k'-th feature extracted from k-th sensor for the j-th hypothesis, received by the fusion center:

$$y_l - \mu_{jkk'} = \phi_k(y_{l-1} - \mu_{jkk'}) + n_l \ , \quad l = 1, \ldots, N_k, \tag{1}$$

where ϕ_k is the AR(1) parameter for the k-th sensor, $\mu_{jkk'}$ is the mean of the k'-th feature for the j-th activity, and n_l is the zero-mean noise with variance $\sigma^2_{jkk'}$. Since the features are modeled as Gaussian,

and given that the AR(1) model is *linear*, the M-ary hypothesis test introduced in Section III is equivalent to the generalized Gaussian problem, using the model in (1).

We denote $\mathbf{m}_i$ and $\boldsymbol{\Sigma}_i, i = 1, 2, \ldots, M$ to be the mean vectors and covariance matrices of the observations under each of the M hypotheses, respectively. For completeness, we recall the density of the multivariate Gaussian random variable, $\mathbf{x} = [x_1, \cdots, x_N]$ is given by:

$$f_X(\mathbf{x}) = \frac{(2\pi)^{-N/2}}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \tag{2}$$

where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix. We denote the number of selected features as $f_{k'}$, where $k' \in \{$NOK, ACC, ECG$\}$, and further denote a generic feature as A. For example, the second feature extracted from the first sensor is denoted A_{1,f_2}. Thus, the mean vector and covariance matrix for the observations for hypothesis H_i for $i = 1, \cdots, M$ are given as:

$$\begin{aligned}
\mathbf{m}_i &= \left[\boldsymbol{\mu}_{iA_{1,f_1}}, \ldots, \boldsymbol{\mu}_{iA_{1,f_{\text{NOK}}}}, \boldsymbol{\mu}_{iA_{2,f_1}}, \ldots, \boldsymbol{\mu}_{iA_{2,f_{\text{ACC}}}}, \boldsymbol{\mu}_{iA_{3,f_1}}, \ldots, \boldsymbol{\mu}_{iA_{3,f_{\text{ECG}}}}\right] \\
\boldsymbol{\Sigma}_i &= \text{diag}\left[\boldsymbol{\Sigma}_i(A_{1,f_1}), \ldots, \boldsymbol{\Sigma}_i(A_{1,f_{\text{NOK}}}), \boldsymbol{\Sigma}_i(A_{2,f_1}), \ldots, \boldsymbol{\Sigma}_i(A_{2,f_{\text{ACC}}}), \boldsymbol{\Sigma}_i(A_{3,f_1}), \ldots, \boldsymbol{\Sigma}_i(A_{3,f_{\text{ECG}}})\right]
\end{aligned} \tag{3}$$

where $\boldsymbol{\mu}_{iA_{k,f_{k'}}}$ and $\boldsymbol{\Sigma}_i(A_{k,f_{k'}})$ are the single-feature statistics. For the k-th sensor, the mean vector is of size $\sum_{k'} N_k f_{k'} \times 1$ and the covariance is of size $(\sum_{k'} N_k f_{k'}) \times (\sum_{k'} N_k f_{k'})$. Thus, the size of the complete covariance matrix, when all features from all sensors are considered, is $(\sum_k \sum_{k'} N_k f_{k'}) \times (\sum_k \sum_{k'} N_k f_{k'})$.

Given the signal model in (1), and incorporating zero-mean channel and measurement noise with variance σ_z^2, the covariance matrix for a particular feature $A_{k,f_{k'}}$ can be expressed as[3]

$$\boldsymbol{\Sigma}(A_{k,f_{k'}}) = \frac{\sigma^2_{A_{k,f_{k'}}}}{1 - \phi^2}\mathbf{T} + \sigma_z^2\mathbf{I}, \tag{4}$$

where $\mathbf{T}$ is a Toeplitz matrix [19] whose first row and column are $[1 \quad \phi \quad \phi^2 \quad \ldots \quad \phi^{N_k-1}]$, and $\mathbf{I}$ is the $N_k \times N_k$ identity matrix. This results in the covariance matrices $\boldsymbol{\Sigma}_j, j = 1, \ldots, M$ being block-Toeplitz matrices. To derive a vector optimization that circumvents an exhaustive search, we may approximate

---

[3]The signal model analysis and derivations that follow adopt a simplified notation; ϕ is used instead of ϕ_k when only a single sensor is being analyzed, and the subscript reintroduced when necessary.

the Toeplitz covariance matrices with their associated circulant covariance matrices[4] given by

$$\mathbf{\Sigma}(A_{k,f_{k'}}) \;=\; \frac{\sigma^2_{A_{k,f_{k'}}}}{1-\phi^2}\mathbf{C} + \sigma^2_z\mathbf{I}, \tag{5}$$

where the matrix $\mathbf{C}$ is a circulant matrix whose first row is identical to that of $\mathbf{T}$.

V. Problem Formulation and Optimization

In this section, we derive the two optimization problems central to this paper: (i) minimizing the transmission cost of the samples from the sensors to the Nokia N95 fusion center, while maintaining a predetermined level of performance defined using the probability of misclassification, and (ii) minimize the probability of error subject to a fixed number of available samples, and independent of the transmission cost. Optimization problem (i) is directly motivated by the need to extend the battery-life of the mobile device, as previously described in Section III-C. Our experiments show that the Nokia N95 is the energy bottleneck, compared to the external ACC and ECG sensors, and thus optimally allocating samples to minimize the total transmission cost achieves this objective. We derive a closed-form approximation for the probability of misclassification in the multihypothesis case via a union bound incorporating the Bhattacharyya coefficients [24] between pairs of hypotheses. Recall that an approximation is derived to circumvent the exhaustive grid search required to compute the optimal allocation of measurements since the fixed total number of samples is an integer; the alternate optimization is continuous-valued and can be solved with significantly lower computational complexity.

A result by Lainiotis [28] provides an upper bound on the probability of misclassification, used to define the performance requirement that must be met while minimizing the transmission power, given as

$$P(\epsilon) \;\leq\; \sum_{i<j}(\pi_i\pi_j)^{1/2}\rho_{ij} \;=\; P_{\text{ub}}(\epsilon), \tag{6}$$

where π_i and π_j are the *a priori* probabilities of hypotheses H_i and H_j from the current state, respectively, and ρ_{ij} is the Bhattacharyya coefficient, which is a measure of the confusability of the two hypotheses,

---

[4]We note that the inverse of the Toeplitz covariance matrix in (4) converges, as N_k increases, to the inverse of the circulant covariance matrix in (5) in the weak sense. Sun et al [46] have derived that a sufficient condition for weak convergence is that the strong norm of the inverse matrices be uniformly bounded, which is the case for the matrix forms in (4) and (5) for $0 < \phi < 1$.

and is canonically defined as [24]

$$\rho_{ij} = \int \sqrt{p_i(\mathbf{x}) p_j(\mathbf{x})} \, d\mathbf{x}, \tag{7}$$

where $p_i(\mathbf{x})$ and $p_j(\mathbf{x})$ are the multivariate densities associated with hypotheses H_i and H_j, respectively. In the case of the multivariate Gaussian, if $p_i(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \mathbf{m}_i, \boldsymbol{\Sigma}_i)$, the Bhattacharyya coefficient is given by [10]:

$$\rho_{ij} = \exp\left(-\left[\frac{1}{8}(\mathbf{m}_i - \mathbf{m}_j)^T \boldsymbol{\Sigma}_h^{-1}(\mathbf{m}_i - \mathbf{m}_j) + \frac{1}{2}\log\frac{\det \boldsymbol{\Sigma}_h}{\sqrt{\det \boldsymbol{\Sigma}_i \cdot \det \boldsymbol{\Sigma}_j}}\right]\right), \tag{8}$$

where $2\boldsymbol{\Sigma}_h = \boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j$, $\mathbf{m}_i$ and $\boldsymbol{\Sigma}_i$ are as defined in (3), and which we rewrite as $\rho_{ij} = \exp(-\psi_{ij})$ for further analysis. The *a priori* probabilities for each of the eight states (1: lie down, 2: sit, 3: sit and fidget, 4: stand, 5: stand and fidget, 6: slow walk, 7: brisk walk, 8: run) are specified in the following estimated transition matrix[5],

$$P = \begin{bmatrix} 0.3 & 0.4 & 0.1 & 0.2 & 0 & 0 & 0 & 0 \\ 0.1 & 0.2 & 0.3 & 0.2 & 0.1 & 0.1 & 0 & 0 \\ 0 & 0.4 & 0.2 & 0.1 & 0.2 & 0.1 & 0 & 0 \\ 0 & 0.1 & 0.1 & 0.3 & 0.2 & 0.2 & 0.1 & 0 \\ 0 & 0.1 & 0.1 & 0.3 & 0.1 & 0.2 & 0.2 & 0 \\ 0 & 0 & 0.1 & 0.1 & 0.1 & 0.3 & 0.2 & 0.2 \\ 0 & 0 & 0 & 0.1 & 0.1 & 0.3 & 0.3 & 0.2 \\ 0 & 0 & 0 & 0.1 & 0.2 & 0.2 & 0.3 & 0.2 \end{bmatrix}, \tag{9}$$

which specifies the probabilities of transitioning from one activity-state to another, and is based on limited free-living datasets[6]. Longer duration free-living deployments are planned to test the KNOWME network, and to develop more robust *a priori* probabilities to be used in the transition matrix. Note that if the current subject state is "Sitting," the upper bound on the probability of misclassification in (6) is computed using six terms since the Sit $\rightarrow$ Brisk Walk and Sit $\rightarrow$ Run transition probabilities are 0. Thus, the optimal allocation depends on both the current state and the possible next state transition probabilities.

Minimizing the probability of misclassification independent of the transmission cost, denoted the

---

[5]The transition matrix is defined as a *right* probability matrix, which is specified as (prior state)$\times$(next state). That is, P_{ij} is the probability of moving from state i to state j, where $i, j \in (1, 2, \ldots, 8)$ for the eight aforementioned activities.

[6]The free-living data used in part to estimate the transition matrix was collected for 20-minute periods at the conclusion of segmented data collection for each subject. The data collection aspect is further described in Section VI-A

December 26, 2010

DRAFT

original optimization, is stated as

$$\min_{\mathbf{N}} \ \log P_{\mathrm{ub}}(\epsilon) \ \text{ subject to} \ \begin{cases} \sum_{k=1}^{K} N_k = N \\ N_k \geq 0 \ \ \forall k \end{cases}, \tag{10}$$

where $\mathbf{N} = (N_1, N_2, \ldots, N_K)$ is the allocation of samples amongst the K sensors.

As described in Section III-C, the internal accelerometer (NOK) measurements are available at a power cost that is significantly less than external Bluetooth sensors measurements (ACC, ECG). We denote the cost of a single measurement from the k-th sensor as C_k, and compute the total transmission power as

$$C_{\mathrm{TX}} = \sum_{k=1}^{K} C_k N_k, \tag{11}$$

where N_k is the number of measurements allocated to the k-th sensor. For the simulations in Section VI-C, and based on our experiments summarized in Section III-C, we assume $C_{\mathrm{NOK}} = 0.585 \cdot C_{\mathrm{ECG}}$ and $C_{\mathrm{ACC}} = 0.776 \cdot C_{\mathrm{ECG}}$. Note that the number of features extracted from the sensor measurements does not affect C_{TX}, since the processing for feature extraction is performed on the mobile device after the samples have been received from all the sensors. The optimization problem considered herein, denoted the *joint optimization*, minimizes this total transmission cost. Thus, the joint optimization problem considered can be stated as

$$\min_{\mathbf{N}} \ C_{\mathrm{TX}} \ \text{ subject to} \ \begin{cases} \log P_{\mathrm{ub}}(\epsilon) \leq \tau \\ \sum_{k=1}^{K} N_k = N \\ N_k \geq 0 \ \ \forall k \end{cases}. \tag{12}$$

A. Low-Complexity Implementation

Given the block-diagonal structure of the covariance matrix in (3), we first decompose the two terms of the argument ψ_{ij} of the Bhattacharyya coefficient for generic hypotheses H_i and H_j, given in (8), as follows:

$$\det \mathbf{\Sigma}_i \ = \ \prod_{k=1}^{K} \prod_{k'=1}^{f_k} \det \mathbf{\Sigma}_i(A_{k,f_{k'}}) \tag{13}$$

$$\Rightarrow \ \log \frac{\det \mathbf{\Sigma}_h}{\sqrt{\det \mathbf{\Sigma}_i \cdot \det \mathbf{\Sigma}_j}} \ = \ \sum_{k=1}^{K} \sum_{k'=1}^{f_k} \log \frac{\det \mathbf{\Sigma}_h(A_{k,f_{k'}})}{\sqrt{\det \mathbf{\Sigma}_i(A_{k,f_{k'}}) \cdot \det \mathbf{\Sigma}_j(A_{k,f_{k'}})}}, \tag{14}$$

and

$$(m_j - m_i)^T \boldsymbol{\Sigma}_h^{-1}(m_j - m_i) = \sum_{k=1}^{K} \sum_{k'=1}^{f_k} (m_{jA_{k,f_{k'}}} - m_{iA_{k,f_{k'}}})^T \boldsymbol{\Sigma}_h^{-1}(A_{k,f_{k'}})(m_{jA_{k,f_{k'}}} - m_{iA_{k,f_{k'}}}). \quad (15)$$

Recall that k iterates over each of the sensors (NOK, ACC and ECG in our case), and k' iterates over the features extracted from each of the sensors. Thus, computing each of the terms for an individual feature $A_{k,f_{k'}}$ (henceforth abbreviated to A_k for the remainder of this section) is sufficient to evaluate the upper bound on the probability of error specified in (6). The structure of the covariance matrix in (3) is block-Toeplitz, or approximated as block-circulant, where the (k,k')-th block, which corresponds to the k'-th single feature from the k-th sensor, is of size $N_k \times N_k$. For every unique allocation of samples amongst sensors, the structure of the covariance matrix is distinct, and thus a combinatorial search over all possible partitions of the total number of samples is required to find the optimal allocation of samples to minimize the probability of error.

To evaluate the term in (13), we use the Toeplitz structure in (4), and rewrite the covariance matrix as follows [20]:

$$\boldsymbol{\Sigma}(A_k) = \boldsymbol{\Sigma}_{\mathrm{D}}(A_k) + \boldsymbol{\Sigma}_{\mathrm{off}}(A_k) = \alpha \mathbf{I} + \frac{\sigma_{A_k}^2}{1 - \phi^2}(\mathbf{T} - \mathbf{I}), \quad (16)$$

where $\alpha = \sigma_{A_k}^2/(1 - \phi^2) + \sigma_z^2$. Given this expansion, the determinant of the covariance matrix can be computed using

$$\det \boldsymbol{\Sigma} = \det \boldsymbol{\Sigma}_{\mathrm{D}} \cdot \det \left(\mathbf{I} + \boldsymbol{\Sigma}_{\mathrm{D}}^{-1} \boldsymbol{\Sigma}_{\mathrm{off}} \right), \quad (17)$$

wherein, using $\mathbf{A} = \boldsymbol{\Sigma}_{\mathrm{D}}^{-1} \boldsymbol{\Sigma}_{\mathrm{off}}$, we now evaluate

$$\det \left(\mathbf{I} + \mathbf{A} \right) = \exp \left(\mathrm{tr} \left(\log \left(\mathbf{I} + \mathbf{A} \right) \right) \right) \quad (18)$$

$$= \exp \left(\mathrm{tr} \left(\mathbf{A} - \frac{\mathbf{A}^2}{2} + \frac{\mathbf{A}^3}{3} - \frac{\mathbf{A}^4}{4} + \cdots \right) \right). \quad (19)$$

The above approximation is valid when $\mathbf{A}$ is a nilpotent matrix [19], which is characterized by both the trace and the determinant of that matrix being equal to zero. For $\mathbf{A} = \boldsymbol{\Sigma}_{\mathrm{D}}^{-1} \boldsymbol{\Sigma}_{\mathrm{off}}$, we see that $\mathrm{tr}(\mathbf{A}) = 0$ (since the diagonal terms of $\boldsymbol{\Sigma}_{\mathrm{off}}$ are all zero), and can compute

$$\det(\mathbf{A}) = \sum_k \beta_k \phi^k \quad \text{where} \quad k \in (N, N+2, N+4, \ldots, 2N-2), \quad (20)$$

for even N^7, where β_k are constants defined in terms of the elements of $\boldsymbol{\Sigma}_\mathrm{D}^{-1}$, which implies $\det(\mathbf{A}) \to 0$ as N gets large since $|\phi| < 1$, thus validating the use of the approximation in (19). For the matrix definitions above, we note that

$$\mathbf{A} = \boldsymbol{\Sigma}_\mathrm{D}^{-1}\boldsymbol{\Sigma}_\mathrm{off} = \frac{1}{\alpha}\frac{\sigma_{A_k}^2}{1-\phi^2}\boldsymbol{\Sigma}_\mathrm{off} = \alpha_0\boldsymbol{\Sigma}_\mathrm{off}, \tag{21}$$

where $\boldsymbol{\Sigma}_\mathrm{off}$ is the matrix explicitly defined in (16), and α_0 is the resulting constant in the above equation. To evaluate the expression in (19), we know $\mathrm{tr}(\mathbf{A}) = 0$, and compute

$$\begin{aligned}
\mathrm{tr}\left(-\frac{\mathbf{A}^2}{2}\right) &= -\mathrm{tr}\left(\frac{\mathbf{A}^2}{2}\right) = -\alpha_0^2\,\mathrm{tr}\left(\frac{\boldsymbol{\Sigma}_\mathrm{off}^2}{2}\right) \\
&= -\alpha_0^2\sum_{k=1}^{N-1}(N-k)\phi^{2k} = \alpha_0^2\sum_{k=1}^{N-1}k[\phi^2]^k - N\alpha_0^2\sum_{k=1}^{N-1}[\phi^2]^k \\
&= \alpha_0^2\left[\phi^2\frac{1-\phi^{2(N-1)}}{(1-\phi^2)^2} - \frac{(N-1)\phi^{2N}}{1-\phi^2}\right] - N\alpha_0^2\left[\phi^2\frac{1-\phi^{2N}}{1-\phi^2}\right],
\end{aligned}$$

where the last simplification is using the following geometric progression identities:

$$\sum_{k=0}^{n}kr^k = r\frac{1-r^n}{(1-r)^2} - \frac{nr^{n+1}}{1-r} \quad \text{and} \quad \sum_{k=0}^{n}r^k = \frac{1-r^{n+1}}{1-r}, \tag{22}$$

which are valid for $r \neq 1$. Thus, using the first two terms of the Taylor expansion, the single feature term in (13) can be approximated as

$$\det\boldsymbol{\Sigma}(A_k) \approx \alpha^{N_k}\exp\left(\alpha_0^2\left[\phi^2\frac{1-\phi^{2(N_k-1)}}{(1-\phi^2)^2} - \frac{(N_k-1)\phi^{2N_k}}{1-\phi^2}\right] - N_k\alpha_0^2\left[\phi^2\frac{1-\phi^{2N_k}}{1-\phi^2}\right]\right), \tag{23}$$

where $\alpha = \sigma_{A_k}^2/(1-\phi^2) + \sigma_z^2$ and α_0 is as defined in (21).

To evaluate the term in (15), the circulant approximation in (5) is employed and we can simplify (15) as

$$(m_{jA_k} - m_{iA_k})^2\sum_{\mathrm{row}=1}^{N_k}\sum_{\mathrm{col}=1}^{N_k}\left[\boldsymbol{\Sigma}_h^{-1}(A_k)\right]_{\mathrm{row,col}}, \tag{24}$$

which shows that we do not need to compute $\boldsymbol{\Sigma}_h^{-1}$, but only require the sum of all the elements of the inverse matrix. To this end, we employ a simple result by Wilansky [52] which states that if the sum of elements in each row of a square matrix is c, then the sum of elements in each row of the inverse is $1/c$.

---

[7] For odd N, the determinant is similar to the form in (20), wherein the smallest power of ϕ in the determinant is $N + 1$.

Note that the sum of the elements of the n-th row of $\boldsymbol{\Sigma}_h(A_k)$ can be simplified as

$$
\begin{aligned}
\text{RowSum}\left[\boldsymbol{\Sigma}_h(A_k)\right] &= \frac{1}{2}\frac{\sigma_{iA_k}^2 + \sigma_{jA_k}^2}{1-\phi^2}(1 + \phi + \phi^2 + \cdots + \phi^{N_k-1}) + \sigma_z^2 \\
&= \frac{1}{2}\frac{\sigma_{iA_k}^2 + \sigma_{jA_k}^2}{1-\phi^2}\frac{1-\phi^{N_k}}{1-\phi} + \sigma_z^2
\end{aligned}
\tag{25}
$$

using the geometric progression in (22). Thus, for a single block of the covariance matrix $\boldsymbol{\Sigma}(A_k)$, we can evaluate the term in (15) as

$$
(m_{jA_k} - m_{iA_k})^2 N_k \left[\frac{1}{2}\frac{\sigma_{iA_k}^2 + \sigma_{jA_k}^2}{1-\phi^2}\frac{1-\phi^{N_k}}{1-\phi} + \sigma_z^2\right]^{-1}.
\tag{26}
$$

Given the per-feature decompositions derived in (14) and (15), the upper bound in (6), for the multi-variate Gaussian case, can be rewritten as

$$
P_{\text{ub}}(\epsilon) = \sum_{i<j} \exp\left[-\sum_{k=1}^{K}\psi_{ij}(N_k) + \frac{1}{2}\log(\pi_i \pi_j)\right],
\tag{27}
$$

where

$$
\psi_{ij}(N_k) = \frac{1}{8}(m_{iA_k} - m_{jA_k})^T \boldsymbol{\Sigma}_h^{-1}(A_k)(m_{iA_k} - m_{jA_k}) + \frac{1}{2}\log\frac{\det\boldsymbol{\Sigma}_h(A_k)}{\sqrt{\det\boldsymbol{\Sigma}_i(A_k)\ \det\boldsymbol{\Sigma}_j(A_k)}},
\tag{28}
$$

where the first term is replaced by (26), and $\det\boldsymbol{\Sigma}(\cdot)$ in the second term is approximated by (23), to yield the continuous-N expression for $P_{\text{ub}}(\epsilon)$

For example, the first order approximation of the expression in (28) can be rewritten, in a simpler functional form, as

$$
\psi_{ij}(N_k) \approx a_{ijk}N_k(1-\phi^{N_k}) + b_{ijk}N_k
\tag{29}
$$

where

$$
a_{ijk} = \frac{1}{16}(m_{iA_k} - m_{jA_k})^2\frac{\sigma_{iA_k}^2 + \sigma_{jA_k}^2}{(1-\phi^2)(1-\phi)}
\tag{30}
$$

$$
\begin{aligned}
\text{and}\ \ b_{ijk} &= \frac{1}{8}(m_{iA_k} - m_{jA_k})^2\sigma_z^2 + \frac{1}{2}\log\left(\frac{1}{2}\frac{\sigma_{iA_k}^2 + \sigma_{jA_k}^2}{1-\phi^2} + \sigma_z^2\right) \\
&\quad -\frac{1}{4}\log\left(\frac{\sigma_{iA_k}^2}{1-\phi^2} + \sigma_z^2\right) - \frac{1}{4}\log\left(\frac{\sigma_{jA_k}^2}{1-\phi^2} + \sigma_z^2\right).
\end{aligned}
\tag{31}
$$

The functional form for the second-order approximation can be similarly derived, but is omitted due to

December 26, 2010

space constraints. We reiterate that the expression in (29) is independent of the discrete block-diagonal structure of the covariance matrix in (3). Thus, a combinatorial search over K integer partitions of N total samples, which requires on the order of $\mathcal{O}(N^{K-1})$ function evaluations, is converted to a continuous-valued vector optimization which is solved with significantly lower complexity, *i.e.* no higher than a polynomial of the number of sensors, $\mathcal{O}(K)$ [5].

B. Lagrangian Optimization

Both the original and joint optimization problems given in (10) and (12), respectively, are convex optimizations problems[8]. For the joint optimization, the associated Lagrangian L is given as

$$L(\mathbf{N}, \lambda_1, \lambda_2) = \sum_{k=1}^{K} C_k N_k + \lambda_1 \left(\log\left[P_{\text{ub}}(\epsilon)\right] - \tau\right) + \lambda_2 \left(\sum_{k=1}^{K} N_k - N\right), \tag{32}$$

where P_{ub} is defined using the approximations in (23) and (26). We note that $L(\mathbf{N}, \lambda_1, \lambda_2)$ is a continuous-valued convex function of $\mathbf{N}$, leads to the following Karush-Kuhn-Tucker (KKT) conditions [5]:

$$C_k + \lambda_1 \frac{\partial}{\partial N_k} \log\left[P_{\text{ub}}(\epsilon)\right] + \lambda_2 = 0, \quad \forall k \tag{33}$$

$$\log\left[P_{\text{ub}}(\epsilon)\right] - \tau = 0, \tag{34}$$

$$\sum_{k=1}^{K} N_k - N = 0, \tag{35}$$

and is solved for the optimal allocation $\mathbf{N}^\star$ and optimal Lagrange multipliers $(\lambda_1^\star, \lambda_2^\star)$, using standard numerical techniques. The optimization is solved for parameters derived using the data from the adolescent test subjects, which is detailed in the next section.

VI. PERFORMANCE ANALYSIS

In this section, we first overview the data collection effort, protocols and test subject details. We then overview a brief statistical analysis of the data, and finally present numerical results for the optimization problems described in Section V, which employ the AR(1)-correlated multivariate Gaussian models.

---

[8]The convexity of the optimization problems is proved by noting that the objective functions and constraints consist of linear/affine functions, which are convex, and the $\log P_{\text{ub}}(\epsilon)$ function. This latter function is of the canonical log-sum-exp form [5], and can also be easily shown to be convex.

A. Data Collection

Until recently, physical activity-monitoring has primarily used commercial accelerometers (see *e.g.* [44]) which require test subjects to first wear the device for an extended period of time followed by data download and post-processing. Standard cut-points (decision regions for single features), which may not be applicable to all demographics, are employed to determine the level of user physical activity. In contrast, KNOWME adopts a different methodology since we use personalized training periods, and offer real-time feedback and user visualization.

Data collection was conducted using one Alive$^{(TM)}$ heart rate monitor [1] and the Nokia$^{(TM)}$ N95 mobile device [2]. A single lead ECG signal is collected with a chest strap on the chest; this heart rate monitor with a built-in accelerometer is positioned in the center of the chest, right below the sternum. The Nokia N95 is placed in a holder on the left hip to record only the accelerometer signal. The optimal sampling algorithm proposed in this work is not restricted to the specific sensors employed, or their respective positions and orientations. However, all sensors must be worn in the *same* location throughout the experiment. Any change in either type or location of sensor used necessitates a new training phase to recalibrate the underlying models. The recognition and optimal sampling algorithms will suffer if location consistency is not maintained.

For each of the three or four sessions, test subjects were required to wear the sensors and perform 8 specific categories of physical activities, for 7 minutes per activity, following a predetermined protocol [45]. The 8 activities performed by each of the test subjects were based on the System for Observing Fitness Instruction Time (SOFIT) protocol [36]. The sequence of the 8 physical activities to be performed was the same for each session for every test subject, unless the test subjects were physically unable to perform the required sequence. The subjects were allowed a rest period between activities for as long as required, and the walking and running portions were at self-selected speeds. The inclusion of the fidgeting component in the sit-and-fidget and stand-and-fidget activities was motivated by it being a component of non-exercise activity thermogenesis (NEAT); fidgeting may constitute a non-trivial source of caloric expenditure and contribute towards combating weight gain [29, 31]. The sitting-and-fidgeting and standing-and-fidgeting activities expend approximately 45% and 70% more energy than standing motionless, respectively [30].

Subject	E1	E2	E3	E4	F1	J1	J2	J3	K1	M1	M2	S1
Gender	F	M	M	F	M	M	F	M	F	F	F	M
Age	16	15	17	14	14	17	13	12	14	17	16	17
BMI %ile	96	99	94	93	97	99	89	97	98	97	91	91
Sessions	4	3	3	4	4	4	4	4	3	3	3	3

TABLE I

DETAILS FOR OVERWEIGHT ADOLESCENT TEST SUBJECTS.

Data from 12 subjects (6 male, 6 female; average age 15.4 ± 1.75 years; average body mass index [9] percentile 95.08 ± 3.40) is reported, wherein 6 subjects performed 4 sessions and 6 subjects performed 3 sessions, on different days and times. Subject details are provided in Table I. The data thus reflects the variability of sensor positions and a variety of environmental and physiological factors.

B. Statistical Analysis of Data

In this section, we analyze data from individual subjects to validate our model assumptions. We first note that the feature selection process, using the SU metric, results in a common set of key features being chosen from both the NOK and ACC sensors; see Figure 6 in Appendix A. The features selected from the mobile-internal and Bluetooth-external accelerometers are (i) median, (ii)-(v) $(20, 40, 60, 80)^{th}$ percentiles, (vi) mean of maxima, and (vii) mean of minima. For the ECG sensor, the key features selected are (i) median of heart-rate and (ii) the noise measure; the HPE coefficients are not selected. Although the aforementioned set of features are dominant, across all sessions and subjects, the optimal allocation for a subject/session is derived using a personalized set of features selected via the SFS algorithm.

1) Gaussianity of Selected Features: Figure 1 show Q-Q plots for the two selected ECG features for the Sitting activity portion of subject J2's first session (upper set of subplots), and for the Standing & Fidgeting activity portion of subject E2's first session (lower set of subplots). These plots are representative of the features that are well modeled as Gaussian, wherein 5-8% of the data points deviate from the normality assumption. However, data corresponding to certain ACC features in some sessions (*e.g.* the percentiles) deviate significantly from normality.

---

[9]The body mass index (BMI = weight/height^2) is a popular measure of relative weight for height and is used by physicians to evaluate the weight status of patients and by epidemiologists to study disease trends in different population samples [40]. Using the standard criteria for overweight and obesity in children and adolescents in the United States, subjects were classified as overweight (age- and gender-specific BMI $\geq$ 85th percentile) or obese (age- and gender-specific BMI $\geq$ 95th percentile) according to CDC growth charts [27].
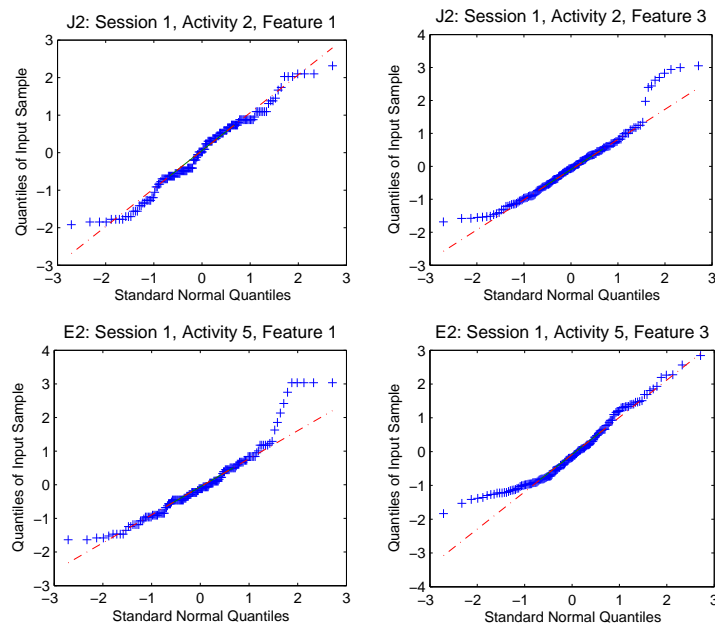
Fig. 1. Q-Q plots for ECG features, for the Sitting and Standing & Fidgeting activities for subjects J2 and E2, validating the Gaussian models employed to develop the optimal allocation.

The correlation coefficient between the two axes of the Q-Q plots is used to test for normality [14, 22], wherein the test statistic is Pearson's correlation coefficient between the ordered observations and the order statistic medians from a normal $\mathcal{N}(0, 1)$ distribution. Since the Q-Q plots for each feature for each of the sensors are generated using several tens of samples, and using the tables in [14], a feature passes the normality test at the 0.05 significance level if the correlation coefficient is greater than 0.987. The features whose Q-Q plots are plotted in Figure 1 each have correlation coefficients greater than 0.9935, and thus pass the normality test, as is evidenced by the Q-Q plots themselves.

Averaged across all 42 sessions from 12 test subjects, 67% of ECG features pass the normality test, and an additional 18% of features have a correlation coefficient of greater than 0.90. The remaining 15% of features have a correlation coefficient of less than 0.90, and are not Gaussian based on the normality test defined in [14]. The percentage of features that are not well modeled as Gaussian increases to approximately 24% in the case of the ACC and the NOK. There does not appear to be a significant correlation between specific activities, types of features or subjects and the Gaussianity of features. However, although the standing and sitting-and-fidgeting activities in the case of the ECG, and the kurtosis, 60th- and 80th-percentile features in the case of the ACC and NOK, appear to deviate most

markedly from Gaussianity, these relationships are not found to be statistically significant (even at the 0.10 significance level).

We note that the approximately 15% of ECG features and 24% of ACC and NOK features, which are clearly not well modeled using the Gaussian distribution, would be more appropriately modeled using other distributions, and considering exponential family distributions as candidates (enabling a low-complexity implementation) is a future avenue of research, but beyond the scope of this paper. Despite the model mismatch for some features, our numerical simulation results validate the energy-savings that are achieved using the optimal sampling algorithm and the low-complexity implementation.

2) Feature-Feature Correlations: An exemplary correlation matrix for the seven features selected using the SU metric from the ACC data, for subject F2's fourth session, is given as:

$$
R_{\text{F1,Session4,ACC}} = \begin{bmatrix} 1.0000 \\ 0.1711 & 1.0000 \\ 0.2667 & 0.1103 & 1.0000 \\ 0.3292 & 0.1448 & 0.2013 & 1.0000 \\ 0.2868 & 0.3940 & 0.1674 & 0.3915 & 1.0000 \\ 0.2874 & 0.2874 & 0.1648 & 0.2486 & 0.4099 & 1.0000 \\ 0.3787 & 0.3080 & 0.3772 & 0.4225 & 0.2612 & 0.2733 & 1.0000 \end{bmatrix} . \tag{36}
$$

We see that all pairs of selected features are not strongly correlated in the above correlation matrix, validating our assumption of independent features given the Gaussian models. Averaged across all 42 sessions from 12 test subjects, 93% of the correlation coefficients in the case of the ACC data are significant at the 0.05 level[10], and the remaining 7% are significant at the 0.10 level. The average feature-feature correlation coefficient for the ACC is 0.2716 ± 0.0439. Correlation coefficients of similar magnitudes and distributions to the ACC data are obtained in the case of the NOK data, wherein 90% of the coefficients are significant at the 0.05 level, and the remaining at the 0.10 level. For the ECG data, as described in Appendix A, the two features selected are the median of the heart-rate and the noise measure. The correlation coefficient between these two features, averaged across all 42 sessions, is 0.3027 ± 0.0672, wherein 94% of the coefficients are significant at the 0.05 level, and the remaining

---

[10]The stated significance level relates to a hypothesis test for testing the statistical significance of the Pearson's correlation coefficient, wherein the two hypotheses are $H_0 : \rho = 0$ and $H_1 : \rho \neq 0$. Refer to [12] for details.

December 26, 2010

at the 0.10 level. Note that using the linear correlation metric for feature selection, instead of the SU metric, results in higher feature-feature correlation values, but does not significantly affect the optimal allocation for a particular scenario.

We note that although at least one correlation coefficient is as high as 0.7627, approximately 97% of the feature-feature correlation coefficients are less than 0.4500 after feature selection using the SU metric. Furthermore, assuming independent features (given our Gaussian models) enables us to derive the low-complexity implementation presented in Section V-A. This assumption yields approximately optimal allocations, but the energy-savings due to the low-complexity implementation (as compared to the exhaustive search) are significant as noted at the end of Section V-A.

C. Performance of Optimal Sampling

We employ data from the 12 overweight adolescent test subjects to characterize the performance of the optimal sampling algorithm derived in the previous sections. Figure 2 plots the optimal allocation (focusing on the 50%-100% allocation range) for the original and joint optimizations, in the upper subplot, averaged across all 42 sessions, for an initial *Sitting* state. The x-axis specifies the original optimization, and the multiplicative factors k for the performance thresholds defined as $\tau = kP_{\text{eq}}$, for the joint optimizations, where τ is defined in (12), and P_{eq} is the probability of error that corresponds to an equal allocation of measurements amongst the NOK, ACC and ECG sensors for a particular subject/session. The discarded sessions, in the lower subplot, are the number of sessions where the minimum probability of error achieved did not meet the performance requirement, and are therefore not included in the computed average allocation. Figure 3 is a similar plot, again focusing on the 50%-100% allocation range, for an initial *Brisk Walking* state.

The difference in the optimal allocation between the original and joint optimizations is evident in both Figures 2 and 3, which show that more NOK samples are allocated in the latter case. A comparison of the two figures suggests that the current state of the subject, as well as the performance requirement, does not significantly affect the optimal allocation, however, a greater number of sessions are unable to meet more stringent performance requirements. These results are elucidated in the following subsections, in addition to analyzing the optimal allocation for a subject's single session, and reiterating the importance of personalized training and model parameters in the KNOWME network.
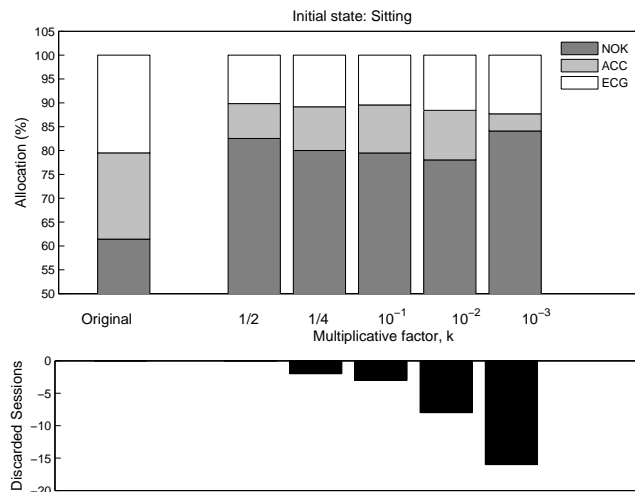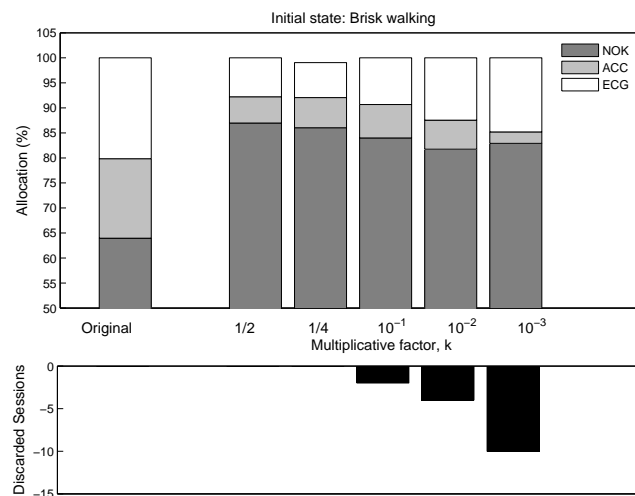
Fig. 2. Optimal allocations, for both the original and joint optimizations, and discarded sessions (joint optimizations wherein the performance requirement was not met) for the Sitting initial state, averaged across all 42 sessions from 12 test subjects. On the x-axis, the multiplicative factor k is specified for the joint optimization.

Fig. 3. Optimal allocations, for both the original and joint optimizations, and discarded sessions (joint optimizations wherein the performance requirement was not met) for the Brisk Walking initial state, averaged across all 42 sessions from 12 test subjects. On the x-axis, the multiplicative factor k is specified for the joint optimization.

1) Varying the $P_{up}(\epsilon)$ threshold, τ: Figures 2 and 3 illustrate the effect of varying the performance requirement in the joint optimization. Given the numerical range of the probability of error achieved for specific subjects and sessions, the threshold is defined as a fraction of the corresponding probability of error that corresponds to equally allocating the measurements, denoted P_{eq}. This is an arbitrary choice and any other threshold could be defined. However, the choice of P_{eq} is analogous to choosing a threshold that

is a function of a system parameter [47], which ensures that the threshold is the same order of magnitude as the computed bound and that the comparison is meaningful. We vary the multiplicative factor k up to three orders of magnitudes, and find that the optimal allocation of samples does not change significantly as the performance requirement becomes more stringent. The noticeable variation in the trend in the case of 10^{-3} can be attributed to the fact that the number of discarded sessions, wherein the performance requirement was not achieved, are more than double than in the 10^{-2} case.

2) Effect of Current Activity: Comparing the optimal allocations achieved in both the original and joint optimizations, for the initial Sitting and Brisk Walking activities in Figures 2 and 3, respectively, suggests that the current activity does not significantly affect the optimal allocation of measurements. This trend is also seen when examining a single subjects' session for each of the eight initial states, and is in contrast to our earlier work [48], wherein the optimal allocation markedly changed based on the initial state. We note that in our previous work, only a subset of states (a maximum of four states) and a single feature were considered, and thus, a change in the initial state had a greater effect on the optimal allocation. In our current work, multiple features are extracted from the sensor measurements, and we consider the pairwise error probabilities between all eight states, which results in the initial state not significantly affecting the optimal allocation of measurements for the exemplary transition matrix, given in (9). However, if an alternative transition matrix were used, as a result of explicitly incorporating the duration of the activities performed, then the optimal allocation may then depend on the initial state. Including time-varying states into our current framework is beyond the scope of this paper, and is a consideration for future work.

3) Single Session Analysis: Having examined the average performance of optimal sampling, we present an exemplary analysis of the algorithm for a single session. Figure 4 shows all possible allocations for $N = 15$ for Session 3 of subject K1; the x-axis is the number of samples allocated to ACC, whereas the multiple lines correspond to different allocations of NOK samples, which are denoted in the legend. Since the total number of samples is constrained, the number of ECG samples is uniquely defined. Note that higher the NOK sample allocation, lower the transmission cost C_{TX}.

The probability of misclassification that corresponds to an equal allocation of samples ($N_1 = N_2 = N_3 = 5$) is plot using a square ($\square$), and the thick black dashed line corresponds to the chosen performance threshold, *i.e.* half the $P(\epsilon)$ for the equal allocation case, denoted P_{eq}. The inverted triangle ($\triangledown$) shows
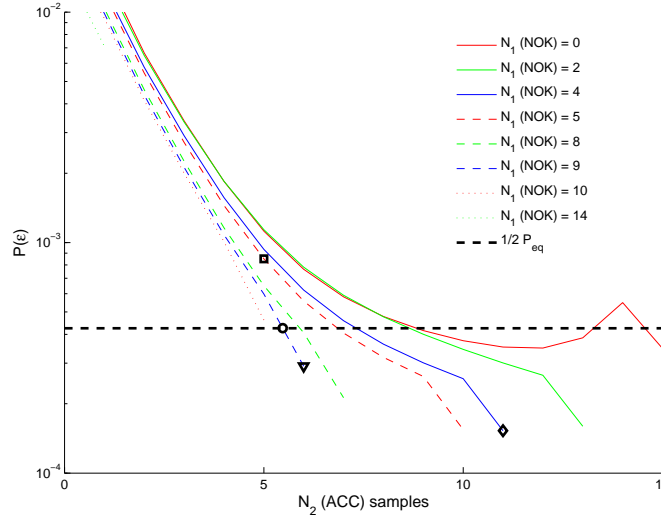
Fig. 4.   Optimal allocation of samples for subject K1's fourth session; equal allocation and optimal allocations for original and joint optimizations are marked.

the optimal allocation of samples for the joint optimization ($N_1^\star = 9, N_2^\star = 6, N_3^\star = 0$) obtained via the exhaustive search, and the circle ($\circ$) shows the corresponding optimal allocation of samples derived as the solution of the low-complexity optimization derived in the previous section. The optimal solution obtained via the exhaustive search corresponds to the allocation $(60\%, 40\%, 0\%)$, while the continuous optimization solution is $(57\%, 40\%, 3\%)$, and is approximately optimal. The diamond ($\diamond$) represents the optimal solution to the original optimization problem, which does not consider the transmission cost.

Figure 4 also illustrates the trade-off between finding an optimal allocation that results in a lower probability of error (for the original optimization) versus one that has a lower total transmission cost (for the joint optimization). As denoted using the diamond, the probability of error is minimum for the allocation ($N_1 = 4, N_2 = 11, N_3 = 0$), but this corresponds to a significantly higher transmission cost since, compared to the optimal allocation that uses 6 Bluetooth ACC samples, 11 Bluetooth ACC samples are used.

4) Energy-Savings from Optimal Sampling: As evidenced in both the single session case, and the averaged performance for both Sitting and Brisk Walking initial states, the optimal solutions to both the original and joint optimization problems provide a measurable advantage over the equal allocation case. Compared to the equal allocation scenario, which uses 33% NOK samples that are available at a significantly lower transmission cost, the solution to the joint optimization problem uses approximately
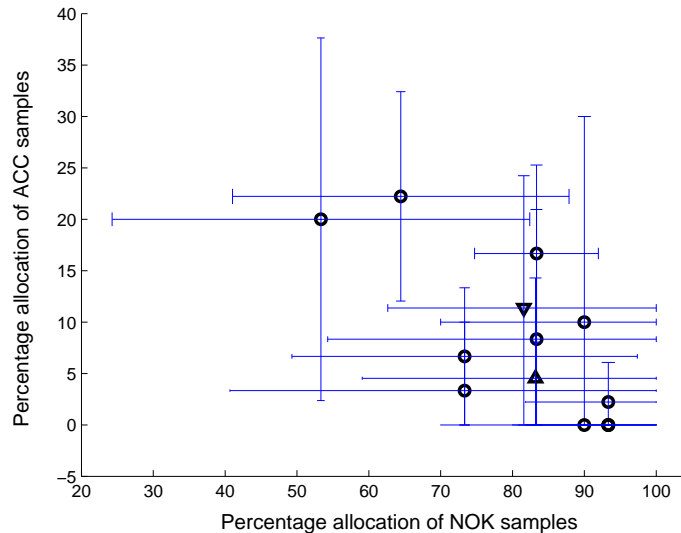
Fig. 5. Optimal allocation of samples for each individual test subject (○), and for subjects grouped by BMI %ile (△ for obese,▽ for overweight), averaged across all available sessions. The error bars are representative of the effect of session variability, and their magnitude reflects that only 3 or 4 sessions were used for all subjects.

75-80% NOK samples, resulting in energy-savings ranging from 18-22%. Similarly, the average energy-savings in the case of the original optimization problem are around 12%. We recall that since the NOK samples are less expensive to obtain than the Bluetooth ACC and ECG samples (see Section III-C), allocating a greater percentage of NOK samples directly results in energy-savings. As is expected, the solution to the joint optimization problem allocates a significantly greater percentage of samples to the NOK, as compared to the original optimization, since the explicit power functional is considered and this results in greater energy-savings.

5) Necessity of Personalized Training: The KNOWME network, and the optimal sampling algorithms developed in this work, employ personalized training, which are used to train the models consequently used for detection and optimal sampling. Figure 5 shows the optimal allocation for individual test subjects, and two instances of test subjects grouped by BMI %ile, wherein each of the data points are averaged over all available sessions for a particular test subject or group of subjects[11]. The x-axis denotes the percentage of allocated NOK samples, and the y-axis denotes the percentage of allocated ACC samples. Bidirectional error bars, which represent the standard deviation of the optimal allocation for that particular

[11]The error bars in Figure 5 are the standard deviation of the allocations and representative of the effect of session variability. Note that the unequal error bars, in the case of some subjects, have been adjusted to reflect that a negative % of samples is not allocated and that the maximum allocation of samples to any particular sensor is 100%.

data point, are included to illustrate the effect of session variability on the optimal allocation. However, since only 3 and 4 sessions are averaged for individual test subjects, the magnitude of the error bars is not entirely unexpected.

The allocation of each of the individual subjects are plotted using ○. Note that the optimal allocation can vary widely between subjects, highlighting the need for personalized training. Furthermore, the optimal allocation for two groups of test subjects are also plotted. Subjects {E3, E4, J2, M2, S1} are classified as overweight (age- and gender-specific BMI ≥ 85th percentile), and subjects {E1, E2, F1, J1, J3, K1, M1} are classified as obese (age-and gender-specific BMI ≥ 95th percentile) according to CDC growth charts using EpiInfo [27]. Optimal allocations for these two groups are plotted using ▽ and △, respectively. The average allocation, for the NOK, ACC and ECG sensors, is 81.6%, 11.4% and 7% for the overweight group, and 83.2%, 4.5% and 12.3% for the obese group, respectively. Thus, although the NOK is preferred for both group due to the lower measurement cost, there is a noticable variation in the allocations between the two groups. This suggests that although grouping by BMI %ile might impact the allocation, it is not the only relevant factor. That is, variables that affect the optimal allocation are not readily identifiable, and thus personalized training may be beneficial for the optimal sampling algorithm.

There exists a noticeable trend amongst the optimal allocation for the test subjects, including the BMI-grouped allocations, averaged over the available sessions. We find that the optimal allocation of samples is normally 70-95% NOK, 25-5% ACC and 5-0% ECG. This trend in the optimal allocations corroborates the result in another of our works [32], which finds that the ECG only minimally contributes, compared to the accelerometer, to accurate detection using SVM and GMM classifiers. However, we note that the range of allocated NOK samples corresponds to a difference in energy savings of up to 10%, and thus conclude that developing personalized models is an effective approach to maximizing the energy-savings.

VII. CONCLUSION AND FUTURE WORK

In this work, we have developed a framework for the optimal allocation of measurements amongst device-internal and Bluetooth-external sensors for activity-detection. A fixed number of measurements are optimized to reduce their transmission cost, while maintaining a performance requirement defined in terms of the probability of detection error. Feature selection is implemented, and correlated Gaussian models are considered. A low-complexity implementation is derived, which significantly reduces the computational burden of deriving the optimal solution, as compared to an exhaustive search. Our models

and algorithms are validated using experimental data from overweight adolescent test subjects, and we find that energy-savings ranging from 18% to 22% are achievable by optimally allocating measurements, compared to the equal allocation scenario. We further find that both the current activity of the subject and the performance requirement of the algorithm do not significantly affect the optimal allocation. However, the optimal allocation varies from subject to subject, and results in variations in energy-savings of up to 10%, suggesting the personalized training approach adopted is an effective method of lengthening the battery-life of the mobile.

Adapting the optimal sampling algorithm for real-time operation is currently being investigated; a variation on the multi-armed bandit formulation, wherein the reward parameters are modeled using the upper bound on the probability of error and the resulting energy-savings from an unequal allocation, is being considered. This formulation can explicitly account for unequal times spent engaged in specific activities, and can account for the power overhead associated with turning on/off the Bluetooth connections at the mobile device. Furthermore, the optimal sampling algorithm is being configured to operate in concert with the higher-dimensional classification algorithms developed in [32]. In particular, an identical set of features can be extracted and selected from optimally allocated samples, to enable robust detection and energy-savings. To this end, a wrapper-based feature selection method [53] would be more appropriate than the filter-based method currently being utilized, since it can be tailored to the specific higher-dimensional classification algorithms being employed. Employing alternative non-Gaussian distributions, especially exponential family distributions, that better model some of the selected features, while ensuring that a low-complexity implementation remains viable, is also being considered.

APPENDIX A

FILTER-BASED FEATURE SELECTION

Feature selection algorithms generally fall into two broad categories [53]. *Wrapper methods* wrap the feature selection around the induction algorithm to be used, using cross-validation to predict the benefits of adding or removing a feature from the feature subset used. *Filter methods* are general preprocessing algorithms that do not rely on any knowledge of the algorithm to be used. The primary advantage of filter methods is their speed and ability to scale to large datasets, and since the optimal allocation does not depend on a specific classifier, we employ sequential forward selection (SFS). Sequential forward selection is a correlation-based feature selection method, wherein the efficacy of a set of features is

evaluated using a heuristic "merit" of a feature subset S containing k features, defined as [16]

$$M_S = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}},\qquad(37)$$

where $\overline{r_{cf}}$ is the mean feature-class correlation,and $\overline{r_{ff}}$ is the average feature-feature inter-correlation. The numerator of (37) can be thought of as providing an indication of how predictive of the class a set of features are; the denominator of how much redundancy there is among the features [16].

We consider a correlation measure based on the information-theoretic metric of *entropy*[12]. We choose the symmetrical uncertainty as a metric, which is defined as [53]

$$SU(X,Y) = 2\left[\frac{I(X;Y)}{H(X) + H(Y)}\right],\qquad(38)$$

where $I(X;Y)$ is the mutual information, and $H(X)$ and $H(Y)$ are the entropies for the random variables X and Y. The SU has values in the range $[0,1]$, with a value of 1 indicating that knowledge of the value of either one completely predicts the value of the other, and a value of 0 indicating independent variables. The SU is computed for every pair of features from all three sensors in order to implement the SFS algorithm.

In the SFS method, each feature that is not already in the current subset is tentatively added to it, and the resulting set of features evaluated using the metric in (37). This evaluation produces a numeric measure of the expected performance of the subset. The effect of adding each feature in turn is quantified by this measure, the best one is chosen, and the procedure continues. This is a standard greedy search procedure, which is guaranteed to find a locally – but not necessarily globally – optimal set of features. The process terminates when the addition of a feature results in the metric not increasing.

The SFS algorithm is implemented for all three sensors. In the case of the NOK and ACC, the best subset of features is selected amongst the 17 total features listed in Box 1. For the ECG, 18 total features are considered; the mean and standard deviation of the inter-peak period, a noise measure, and the first 15 HPE coefficients[13].

Figure 6 shows the percentage of sessions that use each of the features extracted from the NOK, ACC

[12]We have also considered Pearson's correlation coefficient, a linear correlation measure, but do not focus on this metric since it is not always able to capture correlations between variables that are not linear in nature.

[13]We note that the SVM-based detection algorithms in [32] use 60 HPE coefficients, but only 15 of the dominant coefficients are employed to ensure the cardinality of features between the sensors is approximately equal.

and ECG sensors. The average number of features selected by each sensor are NOK: 4.49 ± 1.14, ACC: 4.19 ± 1.40 and ECG: 1.88 ± 0.45. For the SU metric considered, the key features chosen from both the
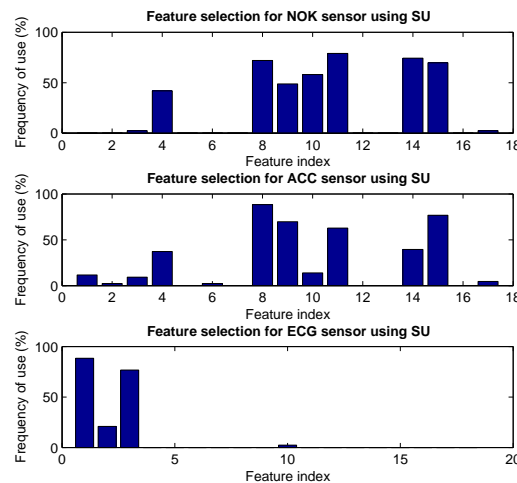

Fig. 6.   Percentage of the total 42 sessions that employ the indexed features, using the SU metric.

inbuilt (NOK) and external Bluetooth (ACC) accelerometers are very similar; the key features selected are (4) median, (8)-(11) $(20, 40, 60, 80)^{th}$ percentiles, (14) mean of maxima, and (15) mean of minima. For the ECG sensor, the key features are (1) median of the heart-rate and (3) the noise measure; the HPE coefficients are not selected.

REFERENCES

[1] Wireless Health Monitors, Alive Technology, 2008. `http://www.alivetec.com`.

[2] Nokia N95, Nokia Corporation, 2008. `http://www.nokiausa.com`.

[3] L. Atallah, B. Lo, et al. Real-time activity classification using ambient and wearable sensors. *IEEE Transactions on Information Technology in Biomedicine*, 13(6):1031–1039, November 2009.

[4] A. Benbasat and J. Paradiso. A framework for the automated generation of power-efficient classifiers for embedded sensor nodes. In *Proc. of SenSys*, pages 219–232, Sydney, Australia, November 2007.

[5] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge Univ. Press, UK, 2004.

[6] Y. Chen, Q. Zhao, et al. Transmission scheduling for optimizing sensor network lifetime: A stochastic shortest path approach. *IEEE Transactions on Signal Processing*, 55(5):2294–2309, May 2007.

[7] L. Collins, S. Murphy, and K. Bierman. A conceptual framework for adaptive preventive interventions. *Prevention Science*, 5(3):185–196, September 2004.

[8] S. Consolvo, D. McDonald, et al. Activity sensing in the wild: A field trial of ubifit garden. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 1797–1806, Florence, Italy, April 2008.

[9] F. Dabiri, H. Noshadi, et al. Light-weight medical bodynets. In *Proceedings of the 2nd International Conference of Body Area Networks*, Article No. 20, Florence, Italy, June 2007.

[10] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. John Wiley, New York, 2001.

[11] P. Dutta, M. Grimmer, et al. Design of a wireless sensor network platform for detecting rare, random and ephemeral events. In *Proceedings of IPSN*, pages 497–502, Los Angeles, CA, April 2005.

[12] A. C. Elliott and W. A. Woodward. *Statistical Analysis Quick Reference Guidebook: With SPSS Examples*. Sage Publications, Los Angeles, CA, 2006.

[13] M. Ermes, J. Parkka, et al. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE Transactions on Information Technology in Biomedicine*, 12(1): 20–26, January 2008.

[14] J. J. Filliben. The probability plot correlation coefficient test for normality. *Technometrics*, 17: 111–117, 1975.

[15] T. Gao, C. Pesto, et al. Wireless medical sensor networks in emergency response: Implementation and pilot results. In *Proceedings of Int'l Conference on Technologies for Homeland Security*, pages 187–192, May 2008.

[16] M. A. Hall. *Correlation-based Feature Selection for Machine Learning*. PhD thesis, The University of Waikato, Hamilton, New Zealand, April 1999.

[17] T. He, B. Krogh, et al. Energy-efficient surveillance system using wireless sensor networks. In *Proceedings of MobiSys*, pages 270–283, Boston, MA, June 2004.

[18] S. V. Hiremath and D. Ding. Evaluation of activity monitors to estimate energy expenditure in manual wheelchair users. In *Proceedings of EMBC'09*, pages 835–838, Minneapolis, MN, September 2009.

[19] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge Univ. Press, Cambridge, UK, 1985.

[20] I. Ipsen and D. Lee. Determinant approximations, 2006. http://www4.ncsu.edu/~ipsen/.

[21] S. Jiang, Y. Cao, et al. Carenet: An integrated wireless sensor networking environment for remote healthcare. In *Proceedings of International Conference on Body Area Networks*, pages 1–3, Tempe, AZ, March 2008.

[22] J. D. Jobson. *Applied Multivariate Data Analysis: Volume 1: Regression and Experimental Design*. Springer, 1991.

[23] E. Jovanov, A. Milenkovic, et al. A wireless body area network of intelligent motion sensors for

computer assisted physical rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 2:6: 16–23, March 2005.

[24] T. Kailath. The divergence and bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15(1):52–60, February 1967.

[25] A. Kalpaxis. Wireless temporal-spatial human mobility analysis using real-time three dimensional acceleration data. In *Proceedings of the International Multi-Conference on Computing in the Global Information Technology (ICCGI)*, pages 1–1, March 2007.

[26] A. Krause, D. Siewiorek, et al. Unsupervised, dynamic identification of physiological and activity context in wearable computing. In *Proceedings of 7th IEEE International Symposium on Wearable Computers*, pages 88–97, White Plains, NY, October 2003.

[27] R. Kuczmarski, C. Ogden, et al. *Vital and Health Statistics Series 11*, chapter 2000 CDC Growth Charts for the United States: Methods and Development. Center for Disease Control, 2002.

[28] D. Lainiotis. A class of upper bounds on probability of error for multihypotheses pattern recognition. *IEEE Transactions on Information Theory*, pages 730–731, November 1969.

[29] J. Levine, N. Eberhardt, and M. Jensen. Role of nonexercise activity thermogenesis in resistance to fat gain in humans. *Science*, 283:212–214, 1999.

[30] J. Levine, S. Schleusner, and M. Jensen. Energy expenditure of nonexercise activity. *American Journal of Clinical Nutrition*, 72:1451–1454, 2000.

[31] J. Levine, E. Melanson, K. Westerterp, and J. Hill. Measurement of the components of nonexercise activity thermogenesis. *American Journal of Physiology – Endocrinology and Metabolism*, 281: E670–675, 2001.

[32] M. Li, V. Rozgic, et al. Multimodal physical activity recognition by fusing temporal and cepstral information. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(4):369–380, August 2010.

[33] T. Linh, S. Osowski, and M. Stodolski. On-line heart beat recognition using Hermite polynomials and neuro-fuzzy network. *IEEE Transactions on Instrumentation and Measurement*, 52(4):1224–1231, 2003.

[34] Y. Liu, B. Veeravalli, and S. Viswanathan. Critical-path based low-energy scheduling algorithms for body area network systems. In *Proc. of the 13th IEEE Int'l Conf. on Embedded and Real-Time Computing Systems and Applications*, pages 301–308, Daegu, Korea, August 2007.

[35] X. Long, B. Yin, and R. M. Aarts. Single-accelerometer-based daily physical activity classification. In *Proc. of EMBC*, pages 6107–6110, Minneapolis, MN, September 2009.

[36] T. McKenzie, J. Sallis, and P. Nader. Sofit: System for observing fitness instruction time. *Journal of Teaching in Physical Education*, 11:195–205, 1991.

[37] M. Quwaider, A. Plummer, et al. Real-time posture detection using body area sensor networks. In *Proceedings of 13th IEEE International Symposium on Wearable Computers (ISWC)*, Linz, Austria, September 2009.

[38] S. Reddy, A. Parker, et al. Image browsing, processing, and clustering for participatory sensing: Lessons from a dietsense prototype. In *Proc. of Workshop on Embedded Networked Sensors*, pages 13–17, Ireland, June 2007.

[39] A. Ross, K. Nandakumar, and A. Jain. *Handbook of Multibiometrics*. Springer, New York, NY, 2006.

[40] T. T. Samaras. *Human Body Size and the Laws of Scaling: Physiological, Performance, Growth, Longevity and Ecological Ramifications*, chapter 2: Human Scaling and the Body Mass Index, pages 17–31. Nova Science Publishers, 2007.

[41] E. S. Sazonov, G. Fulk, et al. Automatic recognition of postures and activities in stroke patients. In *Proc. of EMBC'09*, pages 2200–2203, Minneapolis, MN, September 2009.

[42] E. Shih, P. Bahl, and M. Sinclair. Wake on wireless: an event driven energy saving strategy for battery operated devices. In *Proceedings of MobiCom*, pages 160–171, Atlanta, GA, September 2002.

[43] S.-K. Song, J. Jang, and S.-J. Park. Dynamic activity classification based on automatic adaptation of postural orientation. In *Proceedings of EMBC'09*, pages 6175–6178, Minneapolis, MN, September 2009.

[44] D. Spruijt-Metz, D. Berrigan, et al. *Handbook of Assessment Methods for Obesity and Eating Behaviors, Related Problems and Weight: Measures, Theory and Research*, chapter 6. Sage Publications, Thousand Oaks, CA, 2008.

[45] D. Spruijt-Metz, M. Li, et al. Differentiating physical activity modalities in youth using heartbeat waveform shape and differences between adjacent waveforms. In *Int'l Conf. on Diet and Activity Methods*, Washington DC, 2009.

[46] F.-W. Sun, Y. Jiang, and J. Baras. On the convergence of the inverses of toeplitz matrices and its applications. *IEEE Trans IT*, 49(1):180–190, January 2003.

[47] J. Tasto and W. Rhodes. Noise immunity of threshold decomposition optoelectronic order-statistic filtering. *Optics Letters*, 18(16):1349–1351, 1993.

[48] G. Thatte, M. Li, et al. Energy-efficient multihypothesis activity-detection for health-monitoring

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

34

applications. In *Proc. of EMBC*, pages 4678–4681, Minneapolis, MN, September 2009.

[49] K.-A. Toh, X. Jiang, and W.-Y. Yau. Exploiting global and local decisions for multimodal biometrics verification. *IEEE Transactions on Signal Processing*, 52(10):3059–3072, October 2004.

[50] Y. Wang, J. Lin, et al. A framework of energy efficient mobile sensing for automatic user state recognition. In *Proc. of 7th Int'l Conf. on Mobile Systems, Applications and Services*, pages 179–192, June 2009.

[51] A. Webb. *Statistical Pattern Recognition*. Wiley, 2002.

[52] A. Wilansky. The row-sums of the inverse matrix. *The American Mathematical Monthly*, 58(9): 614–615, November 1951.

[53] L. Yu and H. Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proc. of The Twentieth Int'l Conf. on Machine Leaning (ICML)*, pages 856–863, Washington, D.C., August 2003.