

Energy-Efficient Multihypothesis Activity-Detection for Health-Monitoring Applications

Gautam Thatte, Ming Li, Adar Emken, Urbashi Mitra,
Shri Narayanan, Murali Annavaram and Donna Spruijt-Metz

Abstract—Multihypothesis activity-level detection in a wireless body area network (WBAN) is considered. The fusion center receives biometric samples from heterogeneous sensors. The number of samples collected from each of the sensors is optimized to minimize the probability of misclassification between multiple hypotheses at the fusion center. As each sensor has different discrimination capabilities and the particular Bluetooth-based implemented experimental WBAN, optimal sample allocation results in an overall energy savings. For example, activating all sensors equally yields a 97% accuracy; in contrast, optimal allocation results in 95% accuracy with approximately 30% energy-savings. As the number of samples is an integer, further energy reduction is achieved by developing an unconstrained approximation to the probability of misclassification which allows for a continuous-valued vector optimization. The unconstrained optimization yields approximately optimal allocations with significantly lower complexity.

I. INTRODUCTION

Wireless body area networks (WBANs), an emerging class of sensor networks, are emerging as useful tools for continuous health monitoring. In this paper, we continue to develop the KNOWME network [1], a WBAN that is targeted to applications in pediatric obesity. Herein, we derive an energy-efficient mechanism for activity-detection that can be implemented with low-complexity and examine its efficacy in a “free-living” environment. Our KNOWME WBAN employs heterogeneous sensors in a star topology, which send measurements to a Nokia N95 cellphone fusion center via Bluetooth. The Bluetooth standard for data communication uses a “serve as available” protocol, in which all samples taken by each sensor are collected by the fusion center. However, continuous functioning of Bluetooth requires undesirably high energy consumption, drastically reducing the battery life of the cellphone. There has been significant prior work on activity detection using on-body sensors; however much of the work exploits many accelerometers alone (*e.g.* [3], [7], [9]), or gyroscopes. More recent work [5], [4], [8] using multiple heterogeneous sensors has focused higher layer communication network processing and hardware design. In contrast, our work explicitly focuses on developing the statistical signal processing techniques required for activity-level detection. Previous works on developing energy-efficient WBAN systems have used sleeping/waking cycles [2] and

unused time redistribution [11], among many other techniques, to minimize power consumption. Our work offers a new approach in that the energy-efficiency of the system is a result of an optimized resource allocation (how many samples per sensor?, see Section IV-A). Our contributions are three-fold: development of a signaling/sampling protocol for free-living scenarios using a finite-state machine for transitions of activity (see *e.g.* [3]), energy savings assessment and use of our activity-detection/sampling methods on real data.

II. SYSTEM OVERVIEW

The target functionality of the KNOWME network for activity-level detection is a “free-living” scenario: the sensors and cellphone, worn throughout the day, estimate energy expenditure and serve as a modality for intervention. The operation of the energy-efficient activity-detection mechanism is via alternating training and optimized phases.

During the training phase, the fusion center employs equal allocation for each sensor and detects the current activity/state. The current state is used to derive the optimal allocation which is subsequently employed in the optimized phase wherein a subset of sensors are active. Optimized allocation is dependent on the current activity state and thus is time-varying. The finite-state machine, in Figure 1, is exemplary of the nominal prior probabilities used to determine which sensors to turn off in order to save power, but maintain the detection accuracy. Our previous work [14] considered the optimal allocation problem for the *static* case which did not explicitly account for the current state of the subject. This work develops a framework which results in determining the evolution of the optimal resource allocation as the subject transitions through different states. We observe that the training phase also allows for some self-correction if state detection in the optimized phase is incorrect.

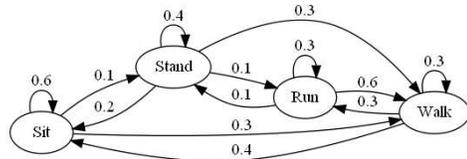


Fig. 1. Finite-state machine with nominal prior probabilities used for the optimal resource allocation case-study.

III. PROBLEM FORMULATION

In this section, we first present the signal model for our wireless body area network, and then outline our optimiza-

This research is supported by the National Center on Minority Health and Health Disparities (supplement to P60 MD002254) and Qualcomm.

G. Thatte et al. are with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089

A. Emken and D. Spruijt-Metz are with the Keck School of Medicine, University of Southern California, Los Angeles, CA 90089

tion problem: minimize the probability of misclassification at the fusion center. We note that obtaining the optimal allocation of samples requires an exhaustive search over the total integer number of samples, N , since all possible partitions of the samples between the heterogeneous sensors must be considered to find the optimal allocation. As the total number of available samples and number of sensors increases, the combinatorial search becomes undesirably computationally expensive. To this end, we develop an analogous optimization problem which yields an approximately optimal solution, but which can be solved using lower complexity continuous-valued vector optimization techniques.

A. Signal Model

Each of the heterogeneous sensors send its biometric samples directly to the cellphone via the Bluetooth protocol. There are a number of key assumptions that we make. First, we approximate the statistics of the features as Gaussian. The features we use are such that one could argue that the central limit theorem holds and thus the Gaussian approximation is not inappropriate. Furthermore, our prior work suggests that the Gaussian model is in close approximation with measured data for the features of interest [1]. To capture the temporal correlation of our features, we use the AR(1) model to represent a biometric time-series, and assume both the sensing and the communication of the measurements to be noisy given the measurement systems and wireless channels, respectively. The AR model has been previously employed to estimate ElectroEncephaloGram (EEG) signals [12] and physiological hand-tremors [16]. Again, we find good approximation with our data and this model. Finally, we presume that the different sensor signals are statistically independent. We have measured the correlation between our features and found the correlation to be low [1]. This assumption is further supported by the fact that certain sensors better discriminate between some subsets of activities than others.

We now propose the following signal model for the decoded and processed samples received by the fusion center:

$$y_i = \theta + z_i, \quad i = 1, \dots, N_k \quad (1)$$

for the k -th sensor, where z_i represents the independent and identically distributed (iid) zero-mean Gaussian measurement and channel noise. For a general feature/sensor A_k , θ is a normally distributed random variable, specified as

$$\theta_j = \mu_{jA_k} + w_i \quad (2)$$

for hypothesis H_j , where w_i represents the *biometric noise* and is modeled using the AR(1) model, *i.e.*

$$w_i = \varphi w_{i-1} + \varepsilon, \quad i = 2, \dots, N_k, \quad (3)$$

for k -th sensor which has been allocated N_k samples, and ε is zero-mean Gaussian with variance $\sigma_{jA_k}^2$. To simplify notation, we omit the hypothesis subscript j when expressions and definitions are applied to a generic hypothesis. We denote the number of samples sent by the K sensors as N_1, N_2, \dots, N_K , respectively, and impose a constraint of N

total samples, *i.e.* $N_1 + N_2 + \dots + N_K = N$, for a specific time-period.

The M -ary hypothesis test using the model in (1) is the generalized Gaussian problem which is specified as

$$H_i: \mathbf{Y} \sim N(\mathbf{m}_i, \mathbf{\Sigma}_i), \quad i = 1, \dots, M, \quad (4)$$

where $\mathbf{m}_i, \mathbf{\Sigma}_i, i = 1, 2, \dots, M$ are the mean vectors and covariance matrices of the observations under the each of the M hypotheses. For the K features A_1, A_2, \dots, A_K from the sensors, the mean vector and covariance matrix for the observations for hypothesis H_j for $j = 1, \dots, M$ are of the form

$$\begin{aligned} \mathbf{m}_j &= [\mu_{jA_1} \ \mu_{jA_2} \ \dots \ \mu_{jA_K}]^T \quad \text{and} \\ \mathbf{\Sigma}_j &= \text{diag}(\mathbf{\Sigma}_j(A_1), \mathbf{\Sigma}_j(A_2), \dots, \mathbf{\Sigma}_j(A_K)), \end{aligned} \quad (5)$$

respectively. Note that μ_{jA_i} is a $N_i \times 1$ vector and $\mathbf{\Sigma}_j(A_i)$ is a $N_i \times N_i$ matrix.

For a particular feature A_k , the covariance matrix can be expressed as

$$\mathbf{\Sigma}(A_k) = \frac{\sigma_{A_k}^2}{1 - \varphi^2} \mathbf{T} + \sigma_z^2 \mathbf{I}, \quad (6)$$

where \mathbf{T} is a Toeplitz matrix whose first row/column is $[1 \ \phi \ \phi^2 \ \dots \ \phi^{N_k-1}]$, and \mathbf{I} is the $N_k \times N_k$ identity matrix. This results in the covariance matrices $\mathbf{\Sigma}_j, j = 1, \dots, M$ being block-Toeplitz matrices. To derive a vector optimization that circumvents an exhaustive search, we may replace the Toeplitz covariance matrices with their associated circulant covariance matrices¹ given by

$$\mathbf{\Sigma}(A_k) \approx \frac{\sigma_{A_k}^2}{1 - \varphi^2} \mathbf{C} + \sigma_z^2 \mathbf{I}, \quad (7)$$

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

B. Probability of Error Derivation

We derive a closed-form approximation for the probability of error in the multi-hypothesis case via a union bound incorporating the Bhattacharyya coefficients between pairs of hypotheses. Extending our work in [14] that considered the static case, the analysis herein depends on the current state \mathcal{S} . A result by Lianiotis [10] provides an upper bound on the probability of error given as

$$P(\epsilon|\mathcal{S}) \leq \sum_{i < j} (P_{i|\mathcal{S}} P_{j|\mathcal{S}})^{1/2} \rho_{ij|\mathcal{S}}, \quad (8)$$

where $P_{i|\mathcal{S}}$ and $P_{j|\mathcal{S}}$ are the *a priori* probabilities for hypotheses H_i and H_j from the current state, and $\rho_{ij|\mathcal{S}}$ is the state-dependent Bhattacharyya coefficient defined as

$$\rho_{ij|\mathcal{S}} = \int [f_i(x) f_j(x)]^{1/2} dx. \quad (9)$$

¹Note the inverse of the Toeplitz covariance matrix in (6) converges to the inverse of the circulant covariance matrix in (7) in the weak sense as the number of samples grows large. A sufficient condition for weak convergence is that the strong norm of the inverse matrices be uniformly bounded [13]. This is the case for our matrix forms for $0 < \phi < 1$.

Thus, the optimization problem considered can be stated as

$$\min_{\mathbf{N}} P(\epsilon|\mathcal{S}) \text{ subject to } \sum_{k=1}^K N_k = N, \quad N_k \geq 0 \quad \forall k, \quad (10)$$

where $\mathbf{N} = (N_1, N_2, \dots, N_K)$ is the allocation of samples amongst the K sensors. In the case of the multivariate Gaussian, if $f_i(x) = N(x; m_i, \Sigma_i)$, the Bhattacharyya coefficient is given by:

$$\rho_{ij|\mathcal{S}} = \exp\left(-\frac{1}{8} m_d^T \Sigma_h^{-1} m_d - \frac{1}{2} \ln \frac{|\Sigma_h|}{\sqrt{|\Sigma_i||\Sigma_j|}}\right), \quad (11)$$

where $m_d = m_i - m_j$, $|\Sigma| = \det \Sigma$, and $2\Sigma_h = \Sigma_i + \Sigma_j$.

Given the block-diagonal structure of the covariance matrix in (5), the terms of the Bhattacharyya coefficient $\rho_{ij|\mathcal{S}}$ are decomposed as follows:

$$\det \Sigma_i = \prod_{k=1}^K \det \Sigma_i(A_k), \quad (12)$$

$$\text{and } \mu_d^T \Sigma_h^{-1} \mu_d = \sum_{k=1}^K \mu_{dA_k}^T \Sigma_h^{-1}(A_k) \mu_{dA_k}, \quad (13)$$

where $\mu_d = \mu_j - \mu_i$ and $\mu_{dA_k} = \mu_{jA_k} - \mu_{iA_k}$. Thus, computing each of the terms for an individual feature A_k is sufficient to evaluate the probability of error specified in (8). For every unique allocation of samples amongst sensors, the structure of the covariance matrix is distinct, and thus an exhaustive 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 (12), we use the Toeplitz structure in (6), and rewrite the covariance matrix as follows [6]:

$$\Sigma(A_k) = \Sigma_D(A_k) + \Sigma_{\text{off}}(A_k) \quad (14)$$

$$= \alpha \mathbf{I} + \frac{\sigma_{A_k}^2}{1 - \phi^2} (\mathbf{T} - \mathbf{I}), \quad (15)$$

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

$$\det \Sigma = \det \Sigma_D \cdot \det (\mathbf{I} + \Sigma_D^{-1} \Sigma_{\text{off}}), \quad (16)$$

wherein, using $\mathbf{A} = \Sigma_D^{-1} \Sigma_{\text{off}}$, we now evaluate

$$\det (\mathbf{I} + \mathbf{A}) = \exp(\text{tr}(\log(\mathbf{I} + \mathbf{A}))) \quad (17)$$

$$= \exp\left(\text{tr}\left(\mathbf{A} - \frac{\mathbf{A}^2}{2} + \frac{\mathbf{A}^3}{3} \dots\right)\right) \quad (18)$$

From the form in (16), and using the geometric progression identity for $\sum_{k=0}^n k r^k$, we evaluate the single feature term in (12) as

$$\det \Sigma(A_k) = \alpha^{N_k} e^{-C[-1 + \phi^{2N_k} - N_k(1 - \phi^{-2})]}, \quad (19)$$

where $C = 1/\alpha^2 (\sigma_{A_k}^2 / (1 - \phi^2))^2 \phi^{-2} / (1 - \phi^{-2})^2$.

To evaluate the term in (13), the circulant approximation in (7) is employed and we can simplify (13) as

$$(\mu_{yA_k} - \mu_{xA_k})^2 \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} [\Sigma_y^{-1}(A_k)]_{ij}, \quad (20)$$

which shows that we only require the sum of all the elements of the inverse matrix. To this end, we employ a result by Wilansky [15] and note that the sum of the elements of the n -th row of $\Sigma(A_k)$ can be simplified as

$$\sum_{j=1}^{N_k} [\Sigma(A_k)]_{nj} = \frac{\sigma_{A_k}^2}{1 - \phi^2} \cdot \frac{1 - \phi^{N_k}}{1 - \phi} + \sigma_z^2 \quad (21)$$

using the geometric progression identity for $\sum_{k=0}^n r^k$. Thus, for a single block of the covariance matrix $\Sigma(A_k)$, we can evaluate the term in (13) as

$$(\mu_{yA_k} - \mu_{xA_k})^2 N_k \left[\frac{\sigma_{yA_k}^2 (1 - \phi^{N_k})}{(1 - \phi^2)(1 - \phi)} + \sigma_z^2 \right]^{-1}. \quad (22)$$

Note that the simplified expressions in (19) and (22) are used to compute the bound on the probability of error in (8), but are independent of the discrete block-diagonal structure of the covariance matrix in (5). Thus, an exhaustive search over K integer partitions of N total samples is converted to a continuous-valued vector optimization which is solved with lower complexity.

IV. NUMERICAL SIMULATIONS

In this section, we present a numerical analysis of the optimal allocation algorithm using experimental data collected using graduate student test subjects. Data from training and testing periods of 10 and 4 minutes, respectively, were collected for seven activities (Sit, Stand, Sit&Fidget, Stand&Fidget, Lie down, Walk and Run) from two accelerometers and an ECG monitor. A 20-minute free-living period concluded the experiments which were conducted in video-monitored simulated living environment. To clarify our results we focus on the two sensor case, although our methods are directly applicable to multiple sensors. Our analysis uses data from two accelerometers and an ECG monitor across three sessions (conducted at different times, on different days) for subject-dependent training, and data from a fourth session is used for testing and yields the detection probabilities reported later in the section. In contrast to our earlier work which considered the static case [14], the numerical results presented here consider the evolution of the optimal allocation as the subject transitions through different activities.

For clarity, we focus on the case with four hypotheses (Sit, Stand, Walk and Run), and two features: the accelerometer variance and the ECG period, which are allocated N_1 and N_2 samples, respectively. The distributions associated with each of the hypotheses for these two features, for a single participant, are shown in Figure 2.

The nominal probabilities used to derive the bound on the probability of error in (8) are as in Figure 1. We evaluate our energy-efficient activity-detection mechanism using a 522-sample (approximately 20 minute) free-living scenario. The testing scenario, decisions taken by the detector and the evolution of the optimal allocation are shown in Figure 3 for the case wherein 10 of every 30 samples is used for training. The third subplot (% ACC) shows the optimal allocation of

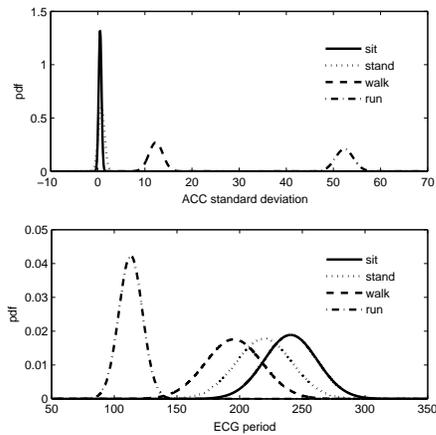


Fig. 2. Gaussian distributions associated with each of the four activities for the ACC StdDev and ECG Period features.

TABLE I
DETECTION ACCURACY FOR DIFFERENT TRAINING SCENARIOS.

Training frequency	Detection accuracy
10 of 50 samples	90.2%
25 of 125 samples	90.4%
10 of 30 samples	95%
25 of 75 samples	92.3%

samples to the accelerometer. We note that allocation is time-varying and depends on the current state of the subject. Table I shows the detection accuracy achieved for different training periods.

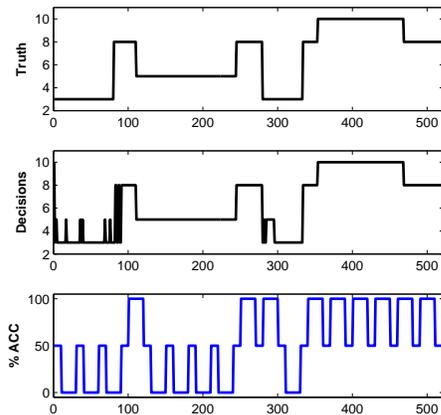


Fig. 3. The free-living testing scenario (3=Sit, 5=Stand, 8=Walk, 10=Run), detector decisions and optimal allocation of samples to ACC.

We use the following algorithm to quantify the energy-savings associated with not using both sensors. Given the probability of error for an optimal allocation of N samples, we compute the minimum $M > N$ equally allocated samples required to achieve the same probability of error. We find that using only the accelerometer or only the ECG results in energy-savings of 43% and 46%, respectively. Thus, we find that as an alternative to 97% accuracy using samples from all sensors, our mechanism achieves 95% accuracy with approximately 30% energy-savings.

A. Complexity Reduction

The vector optimization is significantly lower complexity than the exhaustive search. Given N samples and K sensors, $\mathcal{O}(N^{K-1})$ function evaluations required for an exhaustive search, where a function evaluation is a single computation of the Bhattacharyya coefficient in (11). On the other hand, the continuous-valued vector optimization is independent of the number of total samples available, *i.e.* $\mathcal{O}(1)$.

V. CONCLUSIONS AND FUTURE WORK

An energy-efficient and low-complexity mechanism for free-living activity-detection is developed in this work. We consider activity transitions, and find that unequally allocating samples amongst sensors yields better performance, or equivalent energy-savings, compared to equally allocating samples. For example, we show using real data that activating all sensors equally yields 97% accuracy; in contrast, optimal allocation results in 95% accuracy with approximately 30% energy-savings. The continuous-valued vector optimization derived is lower complexity than an exhaustive search.

We are currently working on the system implementation of the optimal allocation algorithm, and a hidden Markov model approach is being incorporated, similar to the approach in [3], in lieu of the training scenarios. In the future, the KNOWME network will provide excellent tools to accurately measure real-time obesity-related outcomes as well as a modality for intervention.

REFERENCES

- [1] M. Annavaram et al. Multimodal sensing for pediatric obesity applications. In *Proceedings of UrbanSense08*, November 2008.
- [2] A. Benbasat and J. Paradiso. Framework for the automated generation of power-efficient classifiers for embedded sensor nodes. In *Proc. of SenSys*, November 2007.
- [3] S. Biswas and M. Quwaider. Body posture identification using hmm with wearable sensor networks. In *Proc. of BodyNets*, March 2008.
- [4] S. Consolvo et al. Activity sensing in the wild: A field trial of ubifit garden. In *Conf. on Human Factors in Computing Systems*, 2008.
- [5] F. Dabiri, H. Noshadi, H. Hagopian, et al. Light-weight medical bodynets. In *Proc. of BodyNets*, Florence, Italy, June 2007.
- [6] I. Ipsen and D. Lee. Determinant approximations. *Numerical Linear Algebra with Applications*, (Under review), 2006.
- [7] S. Jiang et al. Carenet: An integrated wireless sensor networking environment for remote healthcare. In *Proc. of BodyNets*, March 2008.
- [8] E. Jovanov et al. A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 2:6, March 2005.
- [9] A. Kalpaxis. Wireless temporal-spatial human mobility analysis using real-time three dimensional acceleration data. In *Proc. of Intl. Multi-Conference on Computing in Global Info. Technology*, March 2007.
- [10] D. Lianiotis. A class of upper bounds on probability of error for multihypotheses pattern recognition. *IEEE Trans IT*, November 1969.
- [11] Y. Liu et al. Critical-path based low-energy scheduling algorithms for body area network systems. In *Proc. of RTCSA*, August 2007.
- [12] G. Mohammadi et al. Person identification by using ar model for eeg signals. *Proc. of World Acad. of Sci., Engr. and Tech.*, 11, Feb 2006.
- [13] F.-W. Sun et al. On the convergence of the inverses of toeplitz matrices and its applications. *IEEE Trans IT*, 49(1), January 2003.
- [14] G. Thattai et al. Optimal allocation of time-resources for multihypothesis activity-level detection. In *Proc. of DCOSS*, June 2009.
- [15] A. Wilansky. The row-sums of the inverse matrix. *The American Mathematical Monthly*, 58(9):614-615, November 1951.
- [16] J. Zhang and F. Chu. Real-time modeling and prediction of physiological hand tremor. In *Proc. of ICASSP*, March 2005.