# Power Aware Wireless Data Collection for BSN Data Repositories

Vitali Loseu, Roozbeh Jafari University of Texas at Dallas Embedded Systems and Signal Processing Laboratory Richardson, TX 75080 {vitali.loseu, rjafari}@utdallas.edu

Abstract—Wearable sensor nodes are highly constrained in terms of size, and, as a result, battery size and capacity. During a real time data collection, sensor nodes can communicate data continuously; however, this may reduce the system lifetime. Hence, we suggest an intelligent data collection algorithm that screens the sensor data and transmits only the segments of sensor data that might be of interest. Additionally, the proposed approach does not require extensive system training, since it is based on a flexible Body Sensor Network(BSN) data repository. User can select movements of interest in a repository, and load the sensor nodes with the relevant meta data. Based on the meta data, sensor nodes can decide whether a specific part of the collected sensor data needs to be transmitted. We extend the idea by exploring the idea of dynamically associating the relevance of the data to the amount of energy available at the sensor node.

# I. INTRODUCTION

BSNs are increasingly popular due to their ability to collect detailed and reliable data about various aspects of human life. This property makes BSNs a viable solution in different fields from medical patient monitoring and sports training to gait analyses and fall detection. In general, BSN applications aim to recognize specific observations or specific properties of the desired observations. The most naive solution of this problem is for sensor nodes to forward all of the collected data for future processing to a central basestation. This approach is naive because it is likely to result in sensor nodes communicating data of irrelevant observations. Since wireless communication is the largest factor in energy consumption in BSNs[1], it is more desirable to apply some processing at the local nodes to identify the relevant data before the communication step. This approach allows to reduce the amount of communication in the system without the loss of important information.

It is difficult to guarantee that the approach selected to identify relevant portions of the signal will be absolutely correct all the time. Due to the variations of movements, sensor imprecision, and computational limitations of microcontrollers it is possible that the relevance of a data portion may be assessed incorrectly. As a result, important data sections might be missed, and not communicated to the basetstation. Based on the system training data, however, it is possible to assess the probabilistic properties of misclassification, and define a relationship between the energy expenditure of the system with respect to probability of error. When designing a solution for an application, researchers often customize the software for the specific hardware, existing software tools, and the demands of the application. While this type solution produces effective results, it is not scalable or portable in nature. Finally, this type of approach does not promote data sharing and reuse between different researchers. The shortcomings listed above can be addressed by utilizing the BSN data repository proposed in [2]. In the repository, authors propose to store raw sensor data along with the motion transcripts, and n-grams, extracted from the motion transcripts, to facilitate data mining. We aim to design our data collection utility around the concepts of motion transcripts and n-grams. It will allow for the data collection tool to automatically interface with the BSN repository, making it highly modular and scalable.

In this paper, we propose a power aware opportunistic data collection technique based on the idea of the BSN data repository. We briefly describe the properties of the data repository, and lightweight processing that needs to be executed at individual nodes in real time. Finally, we explore the relationship between the energy expenditure of the system with respect to the probability of missing valuable observations by the system.

### II. DATA REPOSITORY

The repository [2] is designed to extract meta data, or the data used to mine repository, from the signal itself. Intuitively, it is based on the idea of representing BSN sensor readings with text alphabets, and then using fast and light weight techniques, inspired by language processing, to identify portions of the alphabet representation that correspond to important transitions in the signal. This intuition is realized by the following steps. First, when the raw sensor data is collected, the system extracts signal features, such as first and second derivatives. Based on the extracted feature, signal data points are grouped into clusters. The cluster labels are used to define text motion primitives and transcripts, or sequences of motion primitives. Finally, important transitions, or n-grams [3], are extracted from the motion transcripts based on the idea of the information gain. These n-grams are arranged in decision trees for classification decisions. This suggests that in the repository, each movement is represented by a sequence of ngrams, which represent a set of characteristic transitions in the signal. This repository organization is designed to be seamless, fast, computationally inexpensive, and effective. As an initial step of this work, we implemented transcript generation on TelosB nodes using TinyOS to show that such a task can be executed on a sensor node.

## III. PROPOSED APPROACH

Each movement in the repository is represented with a set of n-grams that uniquely identify it. When a user needs to configure a data collection software, they would find the movements of interest in the repository and copy the n-grams of interest onto the appropriate sensor nodes. During the data collection itself, nodes would apply the transcript generation steps to the continuous raw signal data. That would produce a transcript stream representing the raw data. Each sensor would monitor a window of the transcript stream for the n-grams that describe movements of interest. When such n-grams are present in the stream, it suggests that the node is observing one of the movements of interest and should forward the raw data to the base station.

This approach can significantly reduce the energy expenditure of the system, however it faces a potential problem. The system is based on the assumption that each movement can be represented with a set of unique characteristic transitions in the signal. While most of the time this assumption holds, it is possible that due to the variations of movements, sensor imprecision, and computational limitations of microcontrollers movement's observations may appear to have characteristic transitions of other movements. To analyze the potential affect of this phenomenon on the accuracy of the data collection, we designed an experiment that would identify the relationship between energy consumption of the data collection system and the probability of making a mistake.

 TABLE I

 PILOT APPLICATION MOVEMENTS

No.	Description
1	Stand to Sit
2	Sit to Stand
3	Sit to Lie to Sit
4	Bend and Grasp
5	Move Forward (1 step)
6	Move Backward (1 step)

#### **IV. EXPERIMENTAL RESULTS**

For the pilot application, design to identify the relationship between the classification accuracy and energy consumption we selected a system that observes six movements as demonstrated in Table I. We collected the data for the pilot application using TelosB motes with a custom sensor board. The sensor board contains a tri-axis accelerometer, and dual-axis gyroscope. We demonstrate the concept based on the example of the forward facing acceleration of the belt node. In a realistic example, more sensor directions and nodes are used to improve the classification decision, but for the demonstration purposes this is a sufficient example. After the training process, described in Section II, the system created the decision tree in Figure 1. The edges in the figure correspond to n-grams, which identify important transition in the raw sensor readings. The path from the root of the tree to each leaf corresponds to the set of n-grams that uniquely identifies the movement assigned to the leaf node.

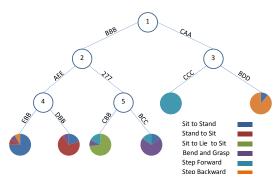


Fig. 1. Decision Tree

To verify the quality of the classification decision individually made by this sensor we applied a cross validation set to the training data, with result listed in Table II. The table shows that the majority of the time the prescribed path produces the correct classification. However, due to error, it is possible that movements might be misclassified. It suggests, that if the data collection system requires a 100% accuracy the system may have to accept and communicate seemingly erroneous movements. For example, in *Stand to Sit* movement, the system would have to communicate the data of *Sit to Stand* and *Bend and Grasp* movements. However, if the system is willing to tolerate an error of 10%, the amount of communication in the system can be cut by a third. This idea is further demonstrated in Figure 2.

TABLE II PILOT APPLICATION ACCURACY

No.	Classified as (%)					
	1.	2.	3.	4.	5.	6.
1. Stand to Sit	82	9	0	9	0	0
2. Sit to Stand	20	70	0	10	0	0
3. Sit to Lie to Sit	0	0	73	9	18	0
4. Bend and Grasp	17	0	0	83	0	0
5. Step Forward	0	0	0	0	100	0
6. Step Backward	11	0	0	0	0	89

# V. CONCLUSION AND FUTURE WORK

In this paper, we proposed an idea of a power aware data collection tool based on the idea of a BSN data repository. We showed that by interfacing sensor nodes with a data repository, and implementing some general, for the repository, signal processing it is possible to create a modular and energy aware data collection tools. We further explored the relationship between the classification error and the energy expenditure of

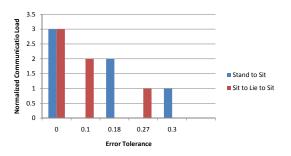


Fig. 2. Classification Accuracy vs Communication Load

such a data collection system on the example of one sensor. In the future, we would like to expand the experiment to include all possible sensors in a BSN. Based on the lessons learned, we will define a dynamic power aware data collection approach that would adjust the acceptable error bound with respect to the amount of energy available at the node.

#### REFERENCES

- V. Raghunathan, C. Schurgers, S. Park, and M. Srivastava, "Energyaware wireless microsensor networks," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 40–50, 2002.
- [2] V. Loseu, H. Ghasemzadeh, L. Khan, and R. Jafari, "A Mining Technique Using N-grams and Motion Transcripts for Body Sensor Network Data Repository," in *Wireless Health*, 2010.
- [3] C. Shannon, "Communication theory of secrecy systems," MD Computing, vol. 15, no. 1, pp. 57–64, 1998.