An Ultra-Low Power Hardware Accelerator Architecture for Wearable Computers Using Dynamic Time Warping

Reza Lotfian and Roozbeh Jafari
The University of Texas at Dallas, Richardson, Texas

Abstract— **Movement monitoring using wearable computers has been widely used in healthcare and wellness applications. To reduce the form factor of wearable nodes which is dominated by battery size, ultra-low power signal processing is crucial. In this paper, we propose an architecture that can be viewed as a hardware accelerator and employs dynamic time warping (DTW) in a hierarchical fashion. The proposed architecture removes events that are not of interest from the signal processing chain as early as possible, deactivating all remaining modules. We consider tunable parameters such as sampling frequency and bit resolution of the incoming sensor readings for DTW to balance the power consumption and classification precision trade-off. We formulate a methodology for determining the optimal set of tunable parameters and provide a solution using Active-set algorithm. We synthesized the architecture using 45nm CMOS and illustrated that a three-tiered module achieves 98% accuracy with a power budget of 1.23µW, while a single level DTW consumes 6.3µW with the same accuracy. We furthermore propose a fast approximation methodology that runs 3200 times faster while introducing less than 3% error over the original optimization for determining the total power consumption.**

Keywords— Hardware Accelerator; Activity Recognition; Granular Decision Making Module; Dynamic Time Warping

I. INTRODUCTION

Continuous human activity recognition and movement monitoring has attracted a significant amount of attention due to potential applications in healthcare. For example, in diagnosis and treatment of neurodegenerative disorders such as Parkinson's disease [1]. Body Sensor Networks (BSNs) facilitate such continuous monitoring. Due to long life-time requirement that can be several days or even months, battery size is the dominant factor which prevents the miniaturization of BSN nodes and limits their wearability. Reducing the power consumption of the nodes will enable creating much smaller sensor nodes and enhancing the comfort for the wearer while satisfying the life-time requirement. An eventual goal is powering the sensor nodes using energy harvesting sources such as body heat or movements which typically provide a power budget of less than 10 µW [2, 3].

To build low power sensor nodes, several studies [4, 5] took advantage of hardware accelerators to perform frequently used tasks in the application more efficiently, *e.g.* filtering, which results in significant power savings compared to using a general purpose microcontroller. Movement monitoring applications have some attractive properties which allow a more aggressive approach with more extensive programmability. These applications are usually looking for a specific or a small set of events or target movements which carry valuable information, such as symptoms of a specific disorder. The notion of wake-up circuitry for identifying less frequent events was introduced in [6]. Granular decision making module for wake-up circuitry was suggested in [7] to recognize body movement using cross-correlation with different bit-resolution. We are using dynamic time warping (DTW) which can be reliably used for movement recognition at very low bit resolutions and a low number of samples in the template [8].

In this paper, data driven architectural optimization method is employed in implementing DTW-based granular decision making module (GDMM) as a hardware accelerator for on-node and real-time movement monitoring. DTW blocks in this module use *bit resolution*, *number of samples* and *acceptance threshold* as tunable parameters. These parameters determine the accuracy of template matching as well as the overall power consumption of the GDMM. Using a numerical optimization method to find optimum tunable parameters, we show how this architecture can reduce the average power consumption in these applications.

II. ACTIVITY RECOGNITION USING GDMM

GDMM is a low power signal pre-conditioning module used prior to more power-hungry processing units such as a microcontroller. We present the proposed architecture in the context of physical movement monitoring where human actions are analyzed by a wearable unit. Our wearable unit is equipped with multiple sensors such as accelerometers and gyroscopes. The sensor readings are continually processed by GDMM and when a movement of interest occurs, the sophisticated processing unit, *e.g.*, a microcontroller will be activated to extract more information about the movements and/or forward the sensor readings to a base station through a wireless link. Since the movement of interest occurs infrequently, the microcontroller remains inactive for most of the time. Our proposed GDMM uses multiple DTW blocks and is optimized to recognize the movement of interest with the minimum energy budget.

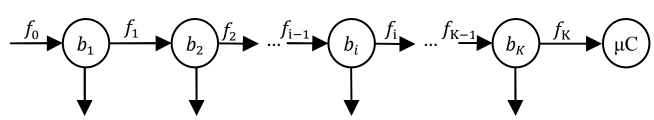

Fig. 1. K-tiered granular decision making module

978-3-9815370-0-0/DATE13/©2013 EDAA

A. Motivation

Data obtained from emerging wearable computers have several properties that are used to enable the proposed novel methodology: events of interest that are intended to be detected using signal processing techniques are often 1) sparse (i.e., the duty cycle on their occurrence is low, typically 0.5-5%, *e.g.*, if a *sit to stand* occurs every 15 minutes with the duration of 5 seconds, its corresponding duty cycle will be 0.5%). The events of interest are 2) slowly changing; and 3) the randomness of the data, even for events that are not of interest, is not significant. That is events are governed by the physics of the human body which constrains the variations. Therefore, non-target movements are easily distinguishable from target movements using template matching blocks with lower complexity (for example lower bit resolution and sampling frequency) operating at a lower power. If non-target movements cannot be identified with less power consuming modules, more sophisticated (and power hungry) template matching blocks are activated.

The notion of processing incoming samples using different sensitivities or accuracies could be expanded to a multi-tiered processing module. Fig. 1 shows a K-tiered decision making module where each block receives a sequence of samples and compares it to a template using DTW algorithm. Each DTW block may observe incoming samples at a lower sampling frequency/bit resolution. The dissimilarity to the template of target movement is determined and compared to a threshold. When the dissimilarity measure is below a threshold, the next block is activated for processing at a higher bit resolution/ sampling frequency. This occurs when the incoming sequence contains a pattern similar to the template of interest. Since each block consists of a DTW template matching and a threshold classifier, the tunable parameters for DTW blocks are the sampling frequency, bit resolution and the acceptance threshold.

Designing a GDMM requires determination of all tunable parameters of every DTW block in the module. There are two metrics for the performance of a module: **power** and **error**. Power is defined as the average energy required to process new incoming sample per time unit. Error is the total number of occurrences of target movements that are misclassified as non-target movements, also known as *false negative error*. Since power and error cannot be described as implicit functions of the tunable parameters, no analytical solution could be found. Therefore, an iterative numerical technique should be employed to find the parameters that satisfy the design requirements. The computational complexity might be a prohibiting factor in any numerical solution to be used. In fact, often the computational complexity will be dominated by the evaluations of power and error functions. In a straightforward solution, for each iteration of the numerical optimization algorithm, exact values of power and error should be determined considering the complete data set (*e.g.*, data obtained during a day). This will significantly prolong the execution of the optimization process, as will be shown in Section IV. Therefore, we propose an approximation technique that considerably reduces the computational complexity and convergence time.

B. Problem Definition

Suppose the first DTW block, b_1, in Fig. 1, receives and processes sequence $S = s_1, s_2, \dots, s_n$ which includes all the sensor readings during a data collection session, for example, sensor readings acquired throughout a day. The DTW block compares S to a template T of size m and generates the minimum warping distance sequence

$$D = d_1, d_2, \dots, d_n = \mathrm{DTW}(S, T) \tag{1}$$

where d_i is the minimum warping distance between the template and the sequence $s_1, s_2, \dots, s_i$ [9]. The finite sequence is defined as a function from the domain $A = \{1,2,\dots,n\}$ to $= \{s_1, s_2, \dots, s_n\}$. In a GDMM, the domain of input sequence of first block is $A_0 = \{1,2,\dots,n\}$. For the other blocks, the domain of input sequence is a subset of A_0 that has been accepted by all previous blocks. A sample is accepted when it potentially belongs to an occurrence of a target movement. To ensure that the whole movement will be captured, for a movement template of size m, we accept a neighborhood of size $2m$ every time a DTW distance below the threshold is detected. This condition is illustrated in (2):

$$\begin{gathered} A_i = \{j | j \in A_{i-1} \\ \exists k \in A_{i-1}, k - \frac{3m}{2} < j < k + \frac{m}{2} : d_k^i < th_i\} \end{gathered} \tag{2}$$

where d_k^i is the distance corresponding to sample k, measured by DTW block i. The sequence of DTW distances for block i becomes: $D_i = DTW_i(S_i, T), S_i : A_i \rightarrow S$.

According to the size of the domain set, the concept of normalized flow is defined as the size of set that is accepted, that is, $f_i = \|A_i\| / \|A_0\|$.

Depending on the bit resolution and the sampling frequency of each DTW block, the power consumption of blocks can be determined. The average energy required to process each sample is called the cost of processing or the cost of a DTW block. This cost, c_i , is obtained by synthesizing hardware of the DTW blocks and performing power analysis using CAD tools. Considering this cost, the total energy for a processing block i becomes:

$$E_i = c_i \|A_{i-1}\| = c_i f_{i-1} \tag{3}$$

The total power consumption of GDMM is:

$$P = F_s \sum_{i=1}^{K} c_i f_{i-1} \tag{4}$$

where F_s is the sampling frequency of the sensors, and K is the number of tiers.

The second metric of performance, error e is defined as the number of target movements that have been rejected incorrectly (false negative). In a K-tiered module, error is:

$$e = \|\{j | j \in Target j \notin A_K\}\| \tag{5}$$

where *Target* is the set of indices for the target movements.

An iterative numerical optimization method should be employed to find the optimum tunable parameters. Since each DTW block is described as a triplet of sampling frequency, bit resolution and threshold, the trial vector of optimization parameters for a K-tiered module is defined as:

$$\boldsymbol{x} = [\boldsymbol{bit}, \boldsymbol{fr}, \boldsymbol{th}] \quad (6)$$

$$\boldsymbol{bit} = [bit_1, \dots, bit_K], \boldsymbol{fr} = [fr_1, \dots, fr_K], \boldsymbol{th} = [th_1, \dots, th_K]$$

Knowing that the error and power are functions of $\boldsymbol{x}$, i.e., $e(\boldsymbol{x})$ and $P(\boldsymbol{x})$, the optimization problem is defined as:

$$\text{Minimize } P(\boldsymbol{x}) = F_s \sum_{i=1}^{K} c_i(\boldsymbol{x}) f_{i-1}(\boldsymbol{x}) \quad \text{such that } e(\boldsymbol{x}) < tolerable\ error \quad (7)$$

C. *Approximate Function Evaluation*

In many engineering optimization problems, the objective and constraint function evaluations require long running times of the computer simulations. Therefore, the computational cost of function evaluations dominates the optimization speed [10]. Here, the evaluation of the objective function $P(\boldsymbol{x})$, subject to the constraint $e(\boldsymbol{x})$, requires the simulation of GDMM using the whole training sequence of samples (*i.e.* data recorded in 24 hours) for *each iteration* and *each block*. To accelerate the process of optimization, we offer the following three modifications to the evaluation function for the error and power in (7):

1. Consider a single DTW block with the cost of c is desired: Several combinations of bit resolutions and sampling frequencies could have the cost of c. For example, DTW blocks with [8 Hz, 6 bit] and [12 Hz, 4 bit] both have 130 nW power costs. Although the combination that has minimum error could be determined implicitly by solving the optimization problem defined in (7), we acquire the best combination prior to the optimization. Having access to the best configurations, *i.e.*, best performing DTW blocks for each power cost, we replace both sampling frequency and bit resolution with cost c, in the trial vector in (6) which simplifies the optimization.
2. The output flow only depends on the threshold and the configuration of the DTW block and it does not depend on the configurations of previous blocks. Since, as we move forward in the signal processing chain, DTW blocks become more selective than their preceding tiers, *i.e.*, $th_i \gg th_{i+1}$, we assume that

$$\nexists i, j : \ d_j^i > th_i \ d_j^{i+1} < th_{i+1} \quad (8)$$

With this assumption, the out-flow of an intermediate DTW block i, with its input set, S_{i-1}, is equal to the flow of the same block with input set S_0. This enables us to estimate the output flow just as a function of the configuration of each block.
3. Since we are only concerned about the false negative error, to evaluate the error function, we only consider the input sequence that carries the target movements. This is a reasonable assumption since we need to ensure that we do not miss the occurrence of a target event. In case of false positives, the only draw-back is the extra cost incurred due to the activation of the microcontroller. This evaluation can be done relatively fast since in the application of movement monitoring, target movements happen quite infrequently.

Using these assumptions, a new trial vector is defined as

$$\bar{\boldsymbol{x}} = [\boldsymbol{c}, \bar{\boldsymbol{f}}], \boldsymbol{c} = [c_1, c_2, \dots, c_K], \bar{\boldsymbol{f}} = [\bar{f}_1, \bar{f}_2, \dots, \bar{f}_K] \quad (9)$$

where $\bar{f}_i$ is the estimated flow for a DTW with cost c_i and threshold of th_i. The optimization problem turns to:

$$\text{Minimize } P(\bar{x}) = F_s \sum_{i=1}^{K} c_i \bar{f}_{i-1} \text{ such that } e(\bar{\boldsymbol{x}}) < tolerable\ error, \ \bar{f}_0 > \bar{f}_1 > \dots > \bar{f}_K > 0 \quad (10)$$

This problem can be solved with numerical derivative-free optimization methods. In this work, we use Active-set method [10] in MATLAB to determine the optimal points.

III. Implementation Methodology

The basic element of GDMM consists of a DTW block and a comparator. The DTW is implemented in RTL, synthesized with TI 45nm standard cell library and the power analysis is done by Synopsys Power Compiler. The operating frequencies for several blocks are reported in Table I. Low leakage cell library with high threshold voltage and low $V_{dd} = 0.67\,V$ is used for power simulations. For each configuration, we set the threshold to accept 5% of total samples and the accuracy is the percentage of the target movements that have been accepted.

We performed an experimental study to assess the effectiveness of DTW blocks in real-world conditions. Ten young subjects, ages 21-35, were asked to perform some movements in an arbitrary fashion for 3 sessions, each 2 hours, while wearing a sensor node on the waist. Each sensor node was equipped with 3-axis accelerometer, 3-axis gyroscope, a microcontroller and a Bluetooth transceiver. The microcontroller was used to control sensors and forward sensor readings to a PC which aggregated sensor readings for offline processing.

IV. Experimental Results

We solved the proposed optimization problem when the error is constrained to be less than 2% and the objective is minimizing the power consumption of the GDMM while considering various movements as the target.

TABLE I. Power Consumption of Different DTW Blocks Implemented Using TI 45nm Standard Cell Library (Vdd=0.67 V)

Configuration		Clock Freq. (Hz)	Cell Area (μ m^2)	Power (*nW*)			Accuracy %
Samp. Freq. (Hz)	Bits			Leakage	Dynamic	Total	
50	**12**	**5000**	**11500**	**275.4**	**24159.3**	**24434.7**	**99.5**
15	**8**	**450**	**5267**	**83.6**	**376.7**	**460.3**	**94.2**
8	**6**	**128**	**1644**	**38.9**	**91.5**	**130.4**	**90.3**
4	**4**	**32**	**633**	**15.4**	**8.8**	**24.2**	**81.8**
2	**2**	**8**	**351**	**8.1**	**1.3**	**9.4**	**41.2**

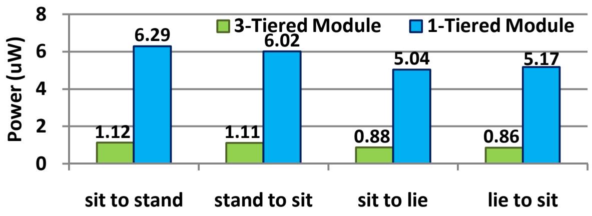

Fig. 2. Comparison of power consumption for 1-tiered module and the optimum 3-tiered module for various activities

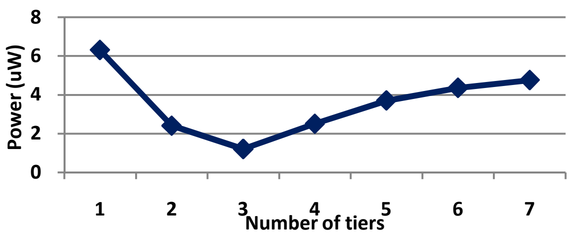

Fig. 3. The minimum power for different number of tiers

Fig. 2 illustrates the average power consumption of the 1-tiered (single DTW block) and 3-tiered (the optimal configuration obtained using the proposed formulation) GDMMs when considering several target movements including *sit to stand*, *stand to sit*, *sit to lie* and *lie to sit*. As shown, using the notion of GDMM (in this example, the 3-tiered architecture) provides over 80% power reduction compared to scenario where GDMM is not used and the signal processing is executed at full bit resolution and sampling frequency (here, 1-tiered DTW block). If we consider Texas Instrument's MSP430F5438 low-power microcontroller executing [16 bit, 50 Hz] DTW operation with 1 MHz clock, the power consumption of the microcontroller measures at 410 µW. Therefore, our proposed GDMM will exhibit a 99.7% power reduction compared to the microcontroller scenario.

Table II shows the activation pattern of tiers of GDMMs optimized for different target movements. As shown, the optimal number of tiers and the activation profile for DTW blocks (or tiers) depends on the target movement.

The effects of the proposed modification in Section II.C on the convergence time of the optimization algorithm are illustrated in Table III. The approximate function evaluation can accelerate iterations of optimization by at least a factor of 3200 while it imposes less than 3% error in estimating the power consumption of GDMM.

We determined the optimal number of tiers using the proposed approximation method, as shown in Figure 3 for the target movement of *sit-to-stand*. The minimum power consumption was obtained when using a 3-tiered GDMM. Since the power is strictly increasing as the number of tiers grows, we chose the 3-tiered module as the global optimal configuration for this movement.

TABLE II. THE OPTIMUM CONFIGURATION OF GDMM FOR DETECTING DIFFERENT MOVEMENTS

Movement	# Occurrences	Optimal # of Tiers	Activation			
			Tier 1	Tier 2	Tier 3	µC
Sit to stand	40	3	1	0.30	0.14	0.005
Stand to sit	40	3	1	0.32	0.14	0.005
Sit to lie	5	3	1	0.28	0.04	0.002
Lie to sit	5	3	1	0.24	0.07	0.002
Bend grasp	3	2	1	0.18	-	0.001

TABLE III. COMPARISON OF ESTIMATED POWER AND ITERATION TIME OF APPROXIMATE AND EXACT OPTIMIZATION METHODS

Architecture	Exact Optimization		Approximate Optimization	
	Power (µW)	Iteration Time (sec)	Power (µW)	Iteration Time (sec)
4-tiered	2.38	720	2.5	0.204
3-tiered	1.17	762	1.23	0.234
2-tiered	2.42	2016	2.43	0.140
1-tiered	6.29	288	6.29	0.052

V. CONCLUSION

In this paper, we introduced an ultra-low power hardware accelerator for the applications of wearable computers. Our proposed architecture, called granular decision making module (GDMM), employs the notion of information-driven tiered signal processing based on dynamic time warping (DTW). The proposed optimization method determines the tunable parameter of DTW blocks of GDMM according to specific characteristics of human motions. Our results provide promising evidence that the proposed architecture will enable the new generation of battery-less wearable computers for the applications of movement monitoring.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation, under grants CNS-1138396 and CNS-1150079. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations.

REFERENCES

[1] D. Gelb *et al*, "Diagnostic criteria for parkinson disease," *Archives of neurology*, vol. 56, no. 1, p. 33, 1999.
[2] L. Swallow *et al*, "A piezoelectric fibre composite based energy harvesting device for potential wearable applications," *Smart Materials and Structures*, vol. 17, no. 2, p. 025017, 2008.
[3] V. Raghunathan *et al*, "Energy-aware wireless microsensor networks," *Signal Processing Magazine, IEEE*, vol. 19, no. 2, pp. 40–50, 2002.
[4] M. Hempstead *et al*, "An ultra low power system architecture for sensor network applications," in *Proc. ISCA'05*, pp. 208–219.
[5] J. Kwong and A. Chandrakasan, "An energy-efficient biomedical signal processing platform," *Solid-State Circuits, IEEE Journal of*, vol. 46, no. 7, pp. 1742–1753, 2011.
[6] B. Calhoun *et al*, "Design considerations for ultra-low energy wireless microsensor nodes," *Computers, IEEE Transactions on*, vol. 54, no. 6, pp. 727–740, 2005.
[7] H. Ghasemzadeh and R. Jafari, "Ultra low power granular decision making using cross correlation: Optimizing bit resolution for template matching," in *Proc.RTAS'11*, pp. 137–146.
[8] R. Jafari and R. Lotfian, "A low power wake-up circuitry based on dynamic time warping for body sensor networks," in *Proc. BSN'11,* pp. 83–88.
[9] Y. Sakurai *et al*, "Stream monitoring under the time warping distance," in *Proc. ICDE 2007* , pp. 1046–1055.
[10] J. Nocedal and S. Wright, *Numerical optimization*. Springer verlag, 1999.