COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 134 (2016) 53–67

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

# An accelerated framework for the classification of biological targets from solid-state micropore data

*Madiha Hanif* [a,b,c], *Abdul Hafeez* [d], *Yusuf Suleman* [a,e], *M. Mustafa Rafique* [f], *Ali R. Butt* [d], *Samir M. Iqbal* [a,b,c,g,h,*]

[a] *Nano-Bio Lab, University of Texas at Arlington, Arlington, TX 76019*
[b] *Department of Bioengineering, University of Texas at Arlington, Arlington, TX 76019*
[c] *Nanotechnology Research Center, University of Texas at Arlington, Arlington, TX 76019*
[d] *Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24060*
[e] *Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX 76019*
[f] *IBM Research, Ireland*
[g] *Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX 76019*
[h] *Department of Urology, University of Texas Southwestern Medical Center at Dallas, Dallas, TX 75390, USA*

ARTICLE INFO

*Article history:*
Received 10 December 2015
Received in revised form
5 May 2016
Accepted 13 June 2016

*Keywords:*
Cancer detection
Pattern detection and classification
Run-time systems
Human cells
Solid-state micropores/nanopores

ABSTRACT

Micro- and nanoscale systems have provided means to detect biological targets, such as DNA, proteins, and human cells, at ultrahigh sensitivity. However, these devices suffer from noise in the raw data, which continues to be significant as newer and devices that are more sensitive produce an increasing amount of data that needs to be analyzed. An important dimension that is often discounted in these systems is the ability to quickly process the measured data for an instant feedback. Realizing and developing algorithms for the accurate detection and classification of biological targets in realtime is vital. Toward this end, we describe a supervised machine-learning approach that records single cell events (pulses), computes useful pulse features, and classifies the future patterns into their respective types, such as cancerous/non-cancerous cells based on the training data. The approach detects cells with an accuracy of 70% from the raw data followed by an accurate classification when larger training sets are employed. The parallel implementation of the algorithm on graphics processing unit (GPU) demonstrates a speedup of three to four folds as compared to a serial implementation on an Intel Core i7 processor. This incredibly efficient GPU system is an effort to streamline the analysis of pulse data in an academic setting. This paper presents for the first time ever, a non-commercial technique using a GPU system for realtime analysis, paired with biological cluster targeting analysis.

© 2016 Elsevier Ireland Ltd. All rights reserved.

* *Corresponding author.* Department of Electrical Engineering, Department of Bioengineering, University of Texas at Arlington, USA; Department of Urology, University of Texas Southwestern Medical Center at Dallas, USA. Tel.: +1 817 272 0228; fax: +1 817 272 7458.
E-mail address: smiqbal@uta.edu (S.M. Iqbal).

http://dx.doi.org/10.1016/j.cmpb.2016.06.001
0169-2607/© 2016 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

Diseases such as cancer can be fully cured, if detected and treated at early stages. The traditional methods like magnetic resonance imaging (MRI) and cytology are intrusive and are not done as part of screening for cancer. The current methods cannot decode the type of cancer in a sample, such as liver cancer or a brain tumor. However, nanoscale biotech devices, for instance nanopores [1,2] and micropores [3,4] enable the translocation of biological targets, such as DNA and human cells in a biological assay at finer granularity. The Coulter counter was patented to detect small particles using resistive-pulse sensing with microscopic tubes, and as the analytes pass through the microchannels, dips in the electrical current are recorded [5–7]. This is the basis of electrical detection in micropores and nanopores. Nanopores provide the ability to detect one molecule at a time [8–17] and the ability to separate individual polymers [18–22]. However, nanometer-scale pores have previously had limitations including, fluctuation in pore currents [23,24], limited residence time of the molecule in the nanopore, nanopore clogging, and poor biomarker selectivity [25,26].

Various computational methods are used to analyze the data attained from nanometer-scale devices, and significant work is focused on the applications of supervised machine-learning algorithms [8,27–36] in order to classify important patterns in gene expression data [37–40]. Simple threshold based on peak-detection algorithms can detect useful patterns in the raw data emerging from ECG and mass spectroscopy [41–43]. A threshold is based on local minimum/maximum, mean, standard deviation, energy or entropy [44–46]. These strategies motivate the design of machine learning approaches for the effective detection and classification of biological targets in the raw data collected from bio-nano sensors.

Nanopores have applications such as rapid detection and characterization of molecules, while micropores are widely used for separating cells [47]. The sensors in this paper are minuscule channels made in thin silicon membranes. Their output is a current signal that is measured in micro- and nano-amperes. Research shows that cancer cells are softer and deform more readily than their healthy counterparts because of their elastic behavior [48]. Such behavior of diseased and healthy cells is recorded as varying patterns (pulses) in the output signal when passed through these devices [49]. The pulses occur at different scales and amplitudes due to the varying size and physical properties of human cells—stiffness and viscosity. Nevertheless, the data collected from such sensors suffer from a large amount of raw data riddled with sensor artifacts. In situ translocation of a characteristic biological assay (0.5 milliliter of a blood sample) results in 10 GB of raw data. The commercial software tools used for the analysis of raw data are limited to smaller datasets, and even a well-trained technician has to spend innumerable hours to process and analyze the data from a simple biological assay. The ability to address an increasing amount of raw data arising from bio-nano devices lies in machine-learning approaches for an effective detection and classification of biological targets in order to accomplish high-quality decision making.

Originally aimed for gaming, graphics processing units (GPUs) have evolved as accelerators into a gamut of compute-intensive scientific applications including bioinformatics and biomedical signal processing [50–68]. The GPU is what translates binary data from the central processing unit (CPU) and converts it into a picture. The tiny dots of an image displayed on a monitor are called pixels. The GPU decides how to use the pixels on a monitor to create an image to reflect the binary data sent to the system. These highly parallel architectures are becoming pervasive toward embarrassingly parallel applications due to their massively parallel architectures. GPUs host clustered cores called streaming processors (SPs), which are further grouped into streaming multiprocessors (SMs). There are varied memory spaces available that range from slower off-chip global memory to the faster on-chip shared memory. On-chip shared memory is faster than off-chip memory, but typically smaller. GPUs are good at doing many tasks at the same time. Programmers optimize memory accesses to global memory either by coalescing memory accesses to global memory or by exploiting the shared memory to reduce the off-chip memory overhead. In parallel, the *k*-nearest neighbor algorithm is used to analyze the data. This technique, in particular, is used because it is a simple analysis tool. However, it is also very effective in this experiment as it delivers results and the necessary classification required for these studies.

This paper describes a novel system-level design that detects pulses from the collected raw data of solid-state micropores, computes useful features of the detected pulses, and finally, classifies unknown samples based on the learned knowledge. The results can be readily used by physicians/scientists to infer useful information for disease diagnosis.

The approach to create a robust and real-time data analysis system included:

- First, a moving-average filtering technique was used to detect pulses from the raw data that stemmed from the diseased and healthy human cells.
- Important features including *width, amplitude, mean, standard deviation, falling slope* and *rising slope* of each pulse, and their statistical significance were computed.
- This enabled instant and reliable classification with the use of *k*-nearest neighbor technique based on the distinguishing features of the detected pulses.
- Finally, parallel *k*-nearest neighbor algorithm was implemented on GPU for biological target cluster detection in order to improve the overall performance of the system.

The implementation of the algorithms for detecting and classifying biological targets from the raw data is described first. The GPU architecture and programming model, as well as the design of the classification algorithms, system components, and experimental setup are discussed in Section 2. Section 3 elaborates the results. Finally, summary and future directions are concluded in Section 4.

## 2. Methods and experimental setup

### 2.1. System overview

The high-level detail of the system modules is provided in Fig. 1. The system is composed of several software compo-

Download English Version:

https://daneshyari.com/en/article/466281

Download Persian Version:

https://daneshyari.com/article/466281

Daneshyari.com