High Speed Processing of Large Amount of Bio-Nano Imaging Data for Cancer Research

G. Wang* and M. Wang**

*Dept. of Biology, Georgia Institute of Technology,
** Dept. of BME, Emory University and Georgia Institute of Technology
maywang@bme.gatech.edu

ABSTRACT

In recent years, Nano science and technology have been researched and used widely, especially opened a new world in bio-related research field. It provided new detection method with high resolution in the tiny space, but also generated huge amount of image data. How to process these data is a challenge problem. In this paper we will present the research work conducted in Bio-MIBLab (Bio-medical Informatics and Bio-imaging Lab) using high performance computer and show some of the preliminary results.

Keywords: quantum dots, nano particles, molecular beacon, high performance, image processing

1 INTRODUCTION

Bionanotechnological development in quantum dots (QDs) and molecular beacons has been used to study cell behavior in vitro and in vivo for disease early diagnosis and therapeutic drug target. These technologies generated huge amount of nanoimaging data that need to be quantified and interpreted for studying cell states. The images usually contain various cellular or sub-cellular environment that are highlighted by multi-color clustered molecular beacons or thousands of single particle - quantum dots. To describe the cell state or to track biomarkers/therapeutic-targets we need to segment cell morphology in 2-D or 3-D, to detect QD colocalization, and to track QDs in 1-D, 2-D, and 3-D. To accomplish this goal, the Bio-MIBLab (Bio-medical Inforamtics and Bioimaging Lab) in Biomedical Engineering of Georgia Tech and Emory University has conducted extensive research in mathematical modeling, computational software, and hardware approaches to analyze large amounts of bionanoimaging data with accuracy and efficiency.

2 IMAGE ANALYSIS

2.1 Sampling and Quantification

The digital image of the experimental result can be represented into 2D matrix form as the following mathematical formula:

$$z = f(x, y)$$

Where (x, y) is the coordinate pair of each pixel, and z is the amplitude. In monochrome, z is the gray level. In color image, z is the intensity mixtures that composed of three different primary colors. The color in our experimental image shows different meaning. The major colors match the primary color of red, and green, such as in **Figure 1**, which represent the color of two types of nano particles. The particle shows generally round shape. If there are aggregations then they are in variant shapes. The background color is black. For each bright spots can be seen from the image, there is possibility of mixture of the different color of particles. So the color of each spot depends on how much overlap occurs at each spot.

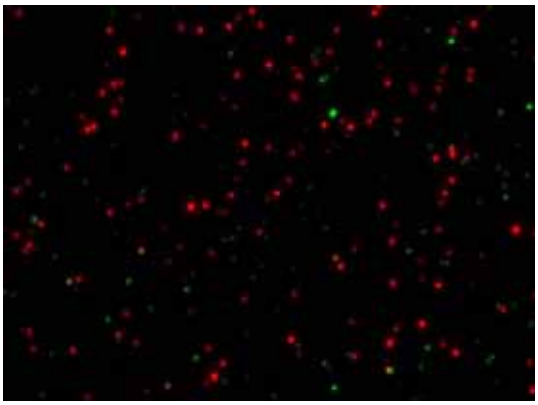


Figure 1

The significance of differentiating the different particles is extremely important because the different binding area of the experiments showed the sensitivity of the nano probes.

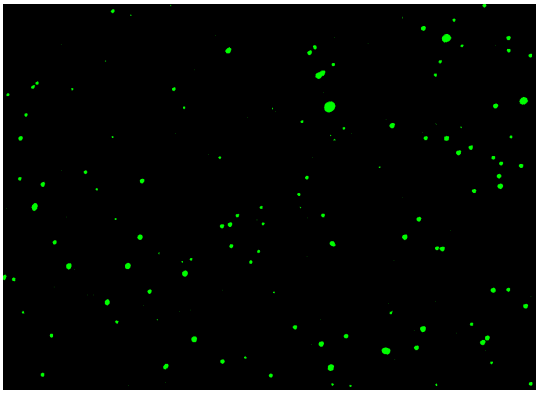


Figure 2

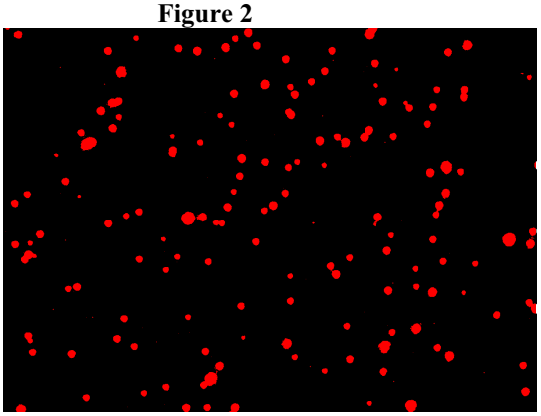


Figure 3

We first sample the whole image in the amplitude domain. The intensity distribution is in a discrete fashion. Then we digitized the amplitude and performed a statistical analysis due to random noise level. The result contributed significantly to the spatial resolution as shown in the Figure 2 and Figure 3. From the zoomed-in result above, we can tell intuitively the corresponding particles overlapped at some region. The **Figure 1, 2 and 3** represent the same segment (1121x849) taken from the original experimental image (3008*2000). **Figure 2** and **3** are processed according the above procedures.

2.2 Classification and Enhancement

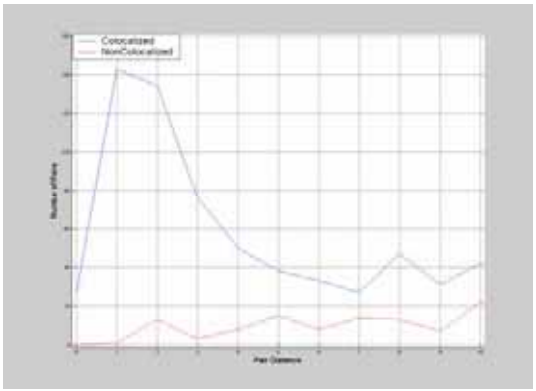
Further more , for cell shape and state quantification, due to the extreme complex cellular environment and heterogeneous cell shapes, we have researched topology-independent algorithms such as adaptive Gradient Field Vector (GVF) snake algorithm and level set techniques, and we have performed a set of numerical simulations are performed repeatedly on every pixel of an image for accuracy. To enhance the image quality, we further utilized grey-level transformation. The transformation mapped the original image pixel value to the processed pixel value. It can be clearly told from the **Figure 2** and **Figure 3**.

2.3 Statistical analysis

The statistical analysis of the result is promising. It provides the proof of our design and hypothesis. The antigen specified nano probes are stick on to the target cells. This can be extended to the cancer cells. The colocalization peak (Figure 4) shown at 1 pixel distance from the experiment compares with control experiment, shows the potential binding site of the green and red nano probes.

Figure 4

3 HIGH PERFORMANCE COMPUTING



The result from above are desirable but the big bottleneck is the processing speed. It cannot be in real-time processed due the huge amount of the data set. Real-time processing is one of our research goals, because tracing the real-time movement of the nano probes is important to define the characteristics of them and for further medical use.

The image size can be as huge as multi-million pixels, and the general purpose central processing unit (CPU) has single execution core and limited memory bandwidth, the processing of data may take a very long time. On tracking 100s to 1000s quantum dots conjugated biomarkers or drug targets in vitro and in vivo for early diagnosis or therapeutics, we have researched various algorithms from simple Gaussian fitting, to robust Kalman tracking. In addition, besides mathematical algorithm design, we have also researched advanced hardware approach. One approach is to use modern programmable graphics processing units (GPUs), which have multiple processing pipelines and wide memory bus, to speed up the computation. This is good for processing data but still cannot meet our real-time requirements.

Because the large amount of dots tracking is very computational intensive, the other approach, we have researched to utilize a state-of-art 68-CPU Itanium2 Linux cluster donated to our lab by Hewlett Packard, to conduct parallel processing and tracking. The characteristics of the images provided us the foundation of utilizing such parallel processing techniques, such as observation field can be divided into segments, only the relative close nano particles have effect on each other.

To prove the concept, we only utilized 12 nodes of Linux cluster. Part of the codes is optimized for the parallel

algorithm. While processing the image with the size 3008x2000 took about 36 hours for a single CPU machine

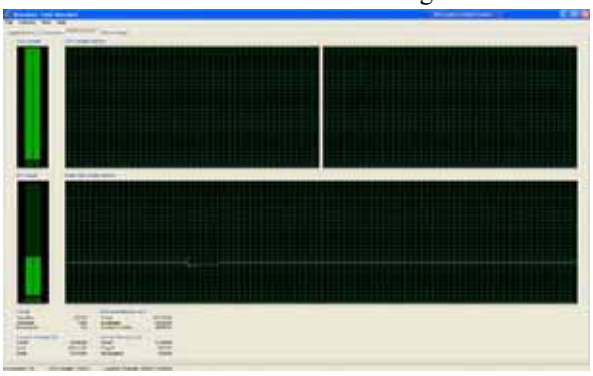


Figure 5

(Figure 5). It was much faster for it to be processed on the HP super computer cluster (Figure 6).

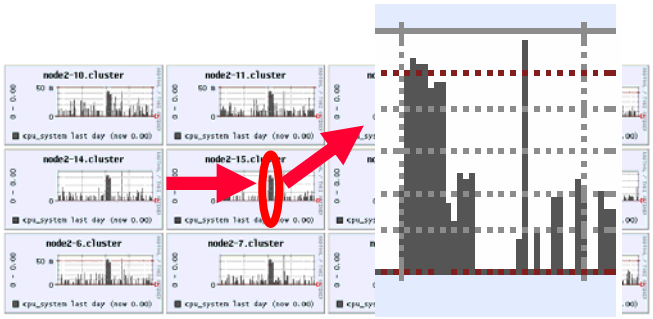


Figure 6

The ideas behind of the parallel high performance can be briefly described by Amdahl's law [2]

$$\text{SpeedUP} = \frac{T(1)}{T(p)} \qquad (1)$$

Where T(p) is the time it takes to execute the program when using p processors and T(1) is the time it takes to execute in serial (one processor).

Ideally, if one processor can execute the program in T(1) time, the p processor should be able to be executed in T(1)/j time. However according the experimental results' shown above, they are not very close to ideal result. Several causes may happen, the efficiency of the hardware, network conditions, overhead of message passing between threads and processes. Also, most important part is the parallelizable of the program. Suppose part of program can be paralleled and part of it can not be paralleled, then the speedup depends on how much percentage of the parallel part to serial part. The Figure 6 of parallel computing using MPI showed speedup pattern**.** The super computer cluster took about more than 4 hours to finish the computation task where the single CPU computer spent more than 36 hours. Because a lot of system time are spent on setting up the communication channel serially at beginning, and the communications between the processes depend on the network status and architectures, so the performance result is not close to the ideal case.

4 CONCLUSION AND FUTURE WORK

The research above showed the promising field of real-time tracking nano-particles with high throughput data. Not only have we shown the achievement of high accuracy image quantification, but also have achieved at least one magnitude of processing speed-up.

In the near future, we will further revise our model and optimize our codes for better efficiency.

Acknowledgement: Thanks for Amit Agwaral in Dr. Shuming Nie's group for providing bionanoimaging data and the high performance computer cluster was donated by HP.

REFERENCES

[1] **Gao X, Cui Y, Levenson RM, Chung LW, Nie S.**, In vivo cancer targeting and imaging with semiconductor quantum dots, Nat Biotechnol. 2004 Aug;22(8):969-76. Epub 2004 Jul 18.

[2]**Amdahl, G.M.** Validity of the single-processor approach to achieving large scale computing capabilities. In *AFIPS Conference Proceedings* vol. 30 (Atlantic City, N.J., Apr. 18-20). AFIPS Press, Reston, Va., 1967, pp. 483-485.