Parallel Computing for Simultaneous Iterative MIMO Tomographic Image Reconstruction by GPUs

[Final Report]

Ricardo Lopez*
Dept of Computer Science
University of Puerto Rico, Rio Piedras
San Juan, PR 00931
ricardo.lopez16@upr.edu

Colleen Rogers†
Dept of Math and Computer Science
Salisbury University
Salisbury, MD 21801
crogers5@gulls.salisbury.edu

Yuanwei Jin
Dept of Engineering and Aviation Sciences
University of Maryland Eastern Shore
yjin@umes.edu

Enyue Lu
Dept of Math and Computer Science
Salisbury University
Salisbury, MD 21801
ealu@salisbury.edu

ABSTRACT

This work is concerned with problem of accelerating inversion algorithms for nonlinear acoustic tomographic imaging by parallel computing on graphics processing units (GPUs). Nonlinear inverse methods for tomographic imaging have many applications. However, these methods often rely on iterative algorithms, thus computationally intensive. In this work, we demonstrate imaging speedup by the Simultaneous Iterative Reconstruction Technique (SIRT) on low cost programmable GPUs. Classic SIRT algorithm has a large memory footprint and slow convergence. In this work, we develop GPU-accelerated SIRT algorithms and implementation strategies. Specifically, we develop scalable algorithms to overcome memory constraints of GPUs. To accelerate convergence, we develop novel weighted SIRT algorithms that utilize non-uniform weights and iterative relaxation factors. We evaluate the performance of our algorithms by the NVDIA Compute Unified Device Architecture (CUDA) programming model on Tesla GPUs. The results show that our algorithms achieve significant speedup in image reconstruction compared with the classic iterative algorithms.

Keywords

Parallel Computing, Tomographic Imaging, GPU Computing

*NSF REU Undergraduate

†NSF REU Undergraduate

1. INTRODUCTION

Full wave based ultrasonic tomographic imaging is able to create high resolution images of the targets of interest and the surrounding inhomogeneous medium. To image, acoustic sources are excited to survey the imaging areas, the scattered wave fields are recorded by sensors placed around the target of interest. Based upon a predetermined wave model, a spatial distribution of the material properties of the medium and the target is created. The imaging process, mathematically, is formulated as an inverse problem, i.e., to infer model parameters $f(\mathbf{r})$ over the imaging region $\mathbf{r} \in \Omega$ from the observed noisy data $y_j = \mathcal{A}_j(f; s_j) + \eta_j$, $j = 1, \cdots, Q$, in response to the j-th excitation source s_j based upon an acoustical wave model denoted by the nonlinear operator $\mathcal{A}_j(\cdot)$. Q is the number of excitation sources and η_j is the noise or disturbance term. Imaging algorithms typically employ inversion methods to reconstruct $f(\mathbf{r})$, for example, Newton's iterative algorithms [5, 7]

$$f^{k+1} = f^k + \omega \delta f^k(s_j) \tag{1}$$

where f^k is the parameter value at the k-th iteration, ω is the relaxiation factor, and $\delta f^k(s_j)$ is the increment parameter value calculated based upon sensor data in response to j-th excitation source s_j, for example, by the propagation and backpropagation method (see our paper [5]) $\delta f^k(s_j) = \left(\mathcal{A}'_j(f^k; s_j)\right)^* (y_j - \mathcal{A}_j(f^k; s_j))$. However, these algorithms are compute intensive and time consuming, which motivates us to develop faster algorithms and explore new methods of employing parallel computing. Contemporary graphics processing units (GPUs) provide economical access to such massively parallel computational capabilities.

In this work, we develop parallel and scalable algorithms implementable on GPUs using the Simultaneous Iterative Reconstruction Technique (SIRT) to address the computational challenges of nonlinear acoustic tomographic imaging problems.

Different than the sequential algorithm (1), also called algebraic reconstruction technique (ART), SIRT is a well known method for image reconstruction following [1, 6]

$$f^{k+1} = f^k + \omega \sum_{m=1}^{M} \delta f^k(S_m) \tag{2}$$

where M is the number of updates that can be calculated simultaneously. The advantages of SIRT are its higher degree of data parallelism, which could lead to faster iterations, and its robustness to noise in measurement data. However, the disadvantages are the higher memory usage $O(M)$ and its slow convergence because more iterations are needed to create an equivalent image compared with ART.

To address the disadvantages of the classic SIRT, we develop GPU-accelerated SIRT algorithms and implementation strategies for nonlinear tomographic imaging. We present three contributions. First, we parallelize the SIRT on GPU for nonlinear inverse algorithm we developed in [5]. Second, we design scalable algorithms to accommodate memory constraints of a given GPU. Third, we develop novel non-uniform weighted SIRT algorithms and demonstrate improved convergence compared with the classic SIRT algorithm.

2. GPU-ACCELERATED SIRT & PERFORMANCE

2.1 Parallelizing SIRT on GPU

We note that two layers of parallelization can be implemented for the iteration equation (2) on GPUs. Residue images $\delta f^k(S_m)$ for m-th sensor group can be calculated in parallel. Furthermore, because we use the 5-point difference equation to calculate the residue images, the computation can be executed in parallel for each spatial grid point of the imaging field at a given time [3, 4, 2]. We utilized NVIDIA's computed unified device architecture (CUDA) programming model, which allowed us to strategically allocate GPU resources (see Table 1) to accelerate the exectution of SIRT by a factor of 65x, compared to the CPU implementation (see Table 2).

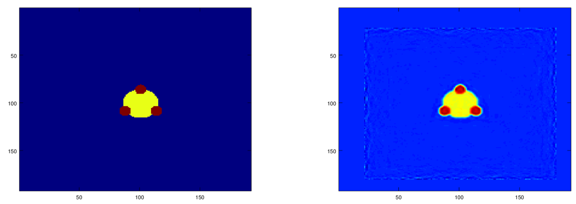

Figure 1: Ground truth image, and SIRT reconstructed image

Table 1: Testing Environment

	GPU	CPU
Device	Tesla k20c	Intel Xeon e5-2260 v2
# of Cores	2496@ 0.71GHz	8 @ 2.20 GHz
RAM	4.69 GB	94.5 GB

2.2 Scalability and Memory Constraints

In the SIRT implementation (2), each of the M image updates must allocate some space in GPU memory in order to execute simultaneously. For large enough M, this can overwhelm the available GPU memory. We scale the memory usage of the SIRT by executing P image updates in parallel, $P \leq M$, and iterating until all M updates have been calculated (see Fig. 2 and Fig. 3).

Table 2: Average SIRT Execution Time (in seconds)

Algorithm	$M = 64$	$M = 32$
GPU	0.7625	0.23898
CPU	46.86984	15.741976
speedup	65.4x	65.9x

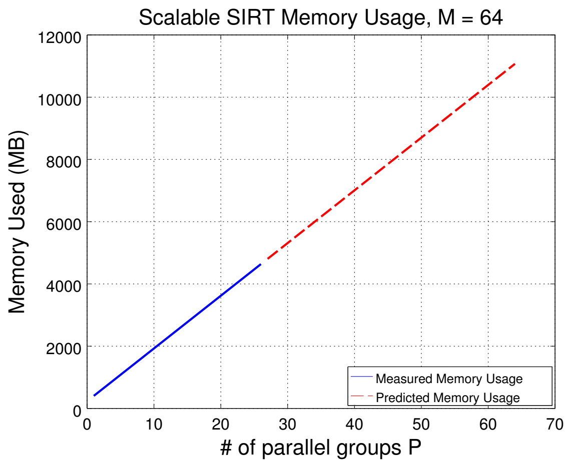

Figure 2: By setting P appropriately, we can control the program's memory usage.

2.3 Non-Uniform Weighted SIRT

Note that in the classic SIRT (2), each intermediate image $\delta f^k(S_m)$ is equally weighted and the relaxation factor ω is a constant. To accelerate the convergence, the weights and the relaxation factor can be modified.

Non-Uniform Weights: We propose to apply non-uniform weights to each $\delta f^k(S_m)$ when we update f, depending on the quality of each $\delta f^k(S_m)$, i.e.,

$$f^{k+1} = f^k + \omega \sum_{m=1}^{M} \frac{1}{d_m} \delta f^k(S_m) \tag{3}$$

where $d_m = \|\delta f^k(S_m) - \overline{\delta f^k}\|$ is the Euclidean norm between the m-th residue image $\delta f^k(S_m)$ and the average image $\overline{\delta f^k} = \frac{1}{M}\sum_{m=1}^{M} \delta f^k(S_m)$.

Iterative Relaxation Factor: We propose an iterative relaxation factor $\omega_i(k)$:

$$f^{k+1} = f^k + \omega_i(k) \sum_{m=1}^{M} \frac{1}{d_m} \delta f^k(S_m) \tag{4}$$

$$w_i(k) = \alpha_i + \frac{\beta_i}{k}, \quad i = k \bmod 2 \tag{5}$$

By using non-uniform weights and the iterative relaxation factor with an optimal α_i and β_i for a given image, we are able to improve the SIRT's convergence (see Fig. 4).

3. CONCLUSIONS

We've shown that we're able to improve the SIRT's convergence in the presence of noise by the use of non-uniform weighting schemes.

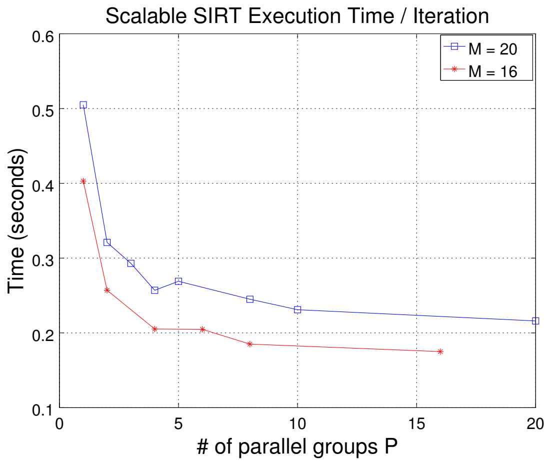

Figure 3: As P increases, execution time of scalable SIRT decreases.

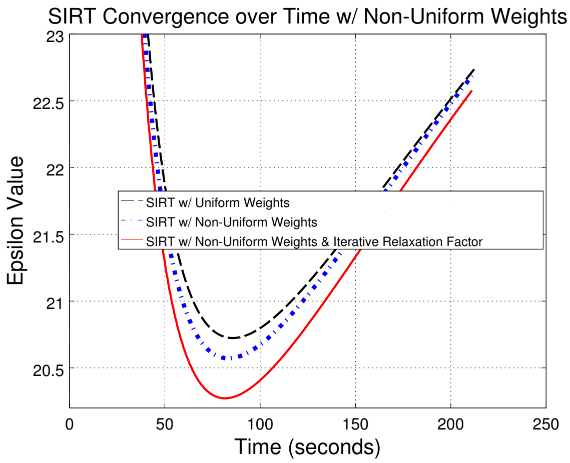

Figure 4: SIRT convergence over time with non-uniform weights.

Although minimal, this SIRT implementation shows improved convergence when compared to the traditional ART, albeit slower (see Fig 5) and with greater memory usage. Given a larger problem set, we conjecture that the Non-Uniform Weighted SIRT will show an even larger improvement over the ART.

Acknowledgments

This work was funded by the National Science Foundation under grant no. CCF-1460900 and the Army Research Office under grant no. W911NF-11-1-0160.

4. REFERENCES

[1] A. H. Andersen and A. C. Kak. Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm. *Ultrasonic Imaging*, 6:81–94, 1984.

[2] P. D. Bello, Y. Jin, and E. Lu. Abstract: GPU Accelerated Ultrasonic Tomography Using Propagation and Backpropagation Method. In *High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion:*, pages 1445–1446, Nov 2012.

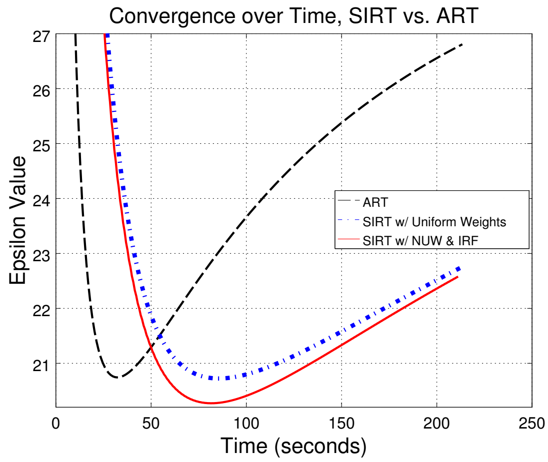

Figure 5: Convergence over time, SIRT vs. ART.

[3] P. D. Bello, Y. Jin, and E. Lu. Gpu accelerated mimo ultrasonic imaging using propagation and back-propagation method. In *IEEE China Summit International Conference on Signal and Information Processing (ChinaSIP)*, pages 374–378, July 2013.

[4] P. D. Bello-Maldonado, A. Rivera-Longoria, M. Idleman, Y. Jin, and E. Lu. Graphics processing units accelerated MIMO tomographic image reconstruction using target sparseness. In *Proc. SPIE*, volume 9109, pages 910900 (1–12), May 2014.

[5] C. Dong and Y. Jin. MIMO nonlinear ultrasonic tomography by propagation and backpropagation method. *IEEE Transactions on Image Processing*, 22(3):1056–1069, March 2013.

[6] B. Keck, H. Hofmann, H. Scherl, M. Kowarschik, and J. Hornegger. Gpu-accelerated SART reconstruction using the CUDA programming environment. In *Proceedings of the SPIE*, volume 7258, pages 72582B (1–12). SPIE, March 2009.

[7] O. Roy, I. Jovanovic, A. Hormati, R. Parhizkar, and M. Vetterli. Sound speed estimation using wave-based ultrasound tomography: Theory and GPU implementation. In *Proceedings of the SPIE: Medical Imaging*, volume 7629, page 76290J, San Diego, CA, 2010. SPIE.