Re-projection of Terabyte-Sized Images

Peter Bajcsy[1], Antoine Vandecreme[1] and Mary C. Brady[1]

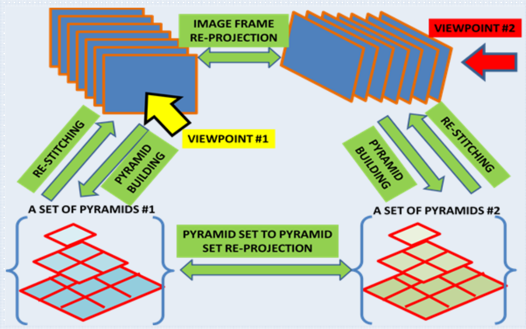
[1]*Software and Systems Division*
Information Technology Laboratory
National Institute of Standards and Technology
Gaithersburg, MD 20899

Abstract: This work addresses the problem of re-projecting a terabyte-sized 3D data set represented as a set of 2D Deep Zoom pyramids. In general, a re-projection for small 3D data sets is executed directly in RAM. However, RAM becomes a limiting factor for terabyte-sized 3D volumes formed by a stack of hundreds of megapixel to gigapixel 2D frames. We have analyzed and benchmarked five methods to perform the re-projection computation in order to overcome the RAM limitation.

How Does One Inspect Terabyte-sized 3D Images From Multiple Viewpoints?

We address the problem of enabling interactive visualization of terabyte-sized 3D images from multiple viewpoints in a web browser.

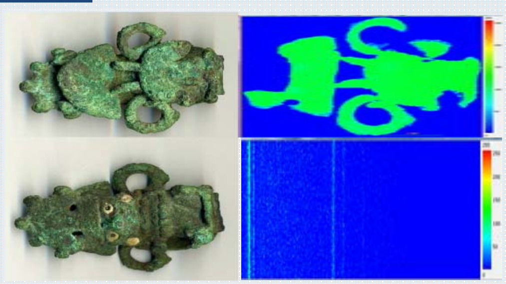
Motivation: With the current limitations of desktop computers in terms of RAM, storage and computation power, the Deep Zoom paradigm [1] has been frequently adopted for visualization of large 2D image data sets. In this work, we focus on re-projecting 3D data sets to deliver multiple 2D views using the Deep Zoom paradigm.



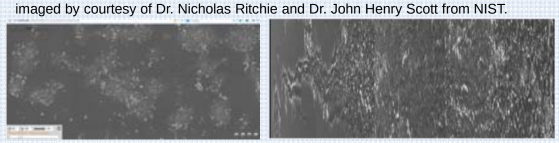
Viewing 3D image volumes from multiple viewpoints using the Deep Zoom pyramids of 2D image cross sections

Approach: Our approach is to utilize the Deep Zoom pyramid representation for large 2D images. Terabyte-sized 3D images are represented as an ordered set of either 2D cross sections or Deep Zoom pyramids. Possible re-projection algorithms for both 3D image representations are analyzed theoretically and experimentally in terms of their computational complexity on a single machine and on a computer cluster with the Hadoop platform.

Test Data

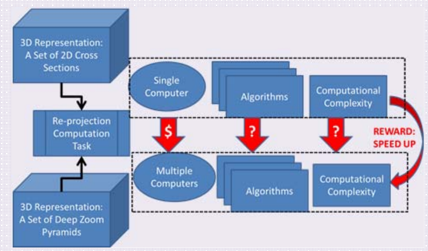


Front and back views of the Moche Cast Figure. Right - Spatial X-Y view (top) and spatio-spectral X-Wavelength view (bottom) of Moche imaged by a scanning electron microscope with energy-dispersive X-ray spectrometry (SEM-EDS). The 3D volume contains 2048 spectral frames with spatial dimensions of 11,520 x 9,984 pixels per frame. The physical sample was provided by courtesy of Prof. Michael Notis, Lehigh University and Prof. Aaron Shugar, Buffalo State College. The physical sample was imaged by courtesy of Dr. Nicholas Ritchie and Dr. John Henry Scott from NIST.



Spatial XY view (left) and spatio-temporal XT view (right) of stem cell colonies imaged every 15 minutes over 5 days using phase contrast microscopy. The 3D XYT volume contains 477 temporal frames with spatial dimensions of 17,866 x 17,193 pixels per frame.

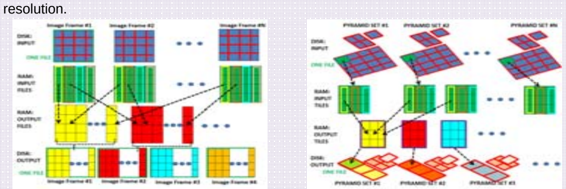
3D Re-Projection Speed-up on Computer Clusters



An overall organization of theoretical analyses and experimental measurements to execute re-projection computations on a single computer and multiple computers.

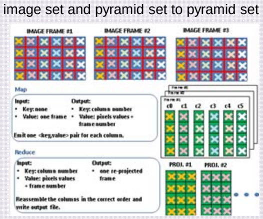
3D Re-projection Algorithms

Single Computer: We compared three implementations to generate pyramids of re-projections. The first one is a re-projection in RAM where the input 3D volume is in RAM and the output 2D frames are generated one after another. The second one is a re-projection based on accessing disk and RAM holding one 2D frame of the input 3D volume and one of the output 3D volume. The re-projection uses file I/O to read-in input 2D frames incrementally and to copy one column/row from input to output. The third one is a re-projection based on copying values from input pyramid tiles to output pyramid tiles where RAM holds only one row of input tiles and one row of re-projected tiles at the highest resolution.

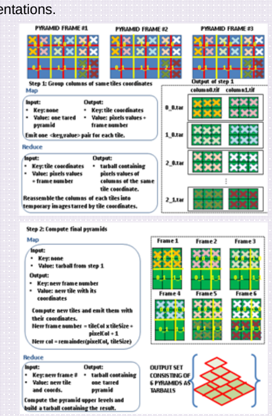


An overview of re-projection computations for an image frame set to an image frame set (left) and a pyramid set to a pyramid set (right). The illustrations focus on RAM memory usage and disk storage.

Multiple Computers: We compared two Map/Reduce algorithms applied to the image set to image set and pyramid set to pyramid set representations.



A description of the Map and Reduce algorithm designed for 3D volume re-projection of image sets on a computer cluster.



A description of the Map and Reduce algorithm designed for 3D volume re-projection of pyramid sets on a computer cluster.

Experimental Hardware

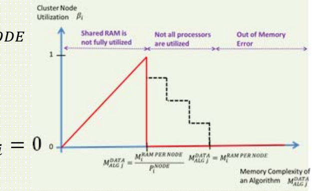
	Specs	Cluster
Hardware	Cluster Nodes	800 computer nodes having from 2 to 16 virtual processors with 4 to 32GB of RAM
	Networking	1Gbit/second
Software	Java Virtual Machine	java version "1.7.0_25" Java(TM) SE Runtime Environment (build 1.7.0_25-b15) Java HotSpot(TM) 64-Bit Server VM
	Hadoop	hadoop-2.0.5-alpha
	Operating System	CentOS 5.9 Linux 2.6.18-274.3.1.el5 x86_64
	File System	Lustre parallel distributed file system

Speed-up: An Extension to Amdahl's Law

$$S(P) = \frac{T(1)}{T(P)} = \frac{1}{\alpha + \frac{1}{P}(1-\alpha)} \qquad P = \sum_{i=1}^{n} \beta_i * P_i^{NODE}$$

$$\textbf{If} \left(\frac{M_i^{RAM\ PER\ NODE}}{P_i^{NODE}} - M_{ALG\ j}^{DATA\ PER\ NODE} \right) > 0$$

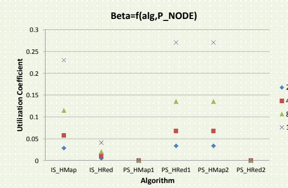
$$\textbf{Then} \quad \beta_i = \frac{P_i^{NODE} * M_{ALG\ j}^{DATA\ PER\ NODE}}{M_i^{RAM\ PER\ NODE}} \qquad \textbf{Else} \ \beta_i = 0$$

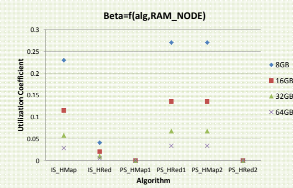


$$ALG^{RAM\ OPTIMAL} = \underset{ALG\ j}{\operatorname{argmin}}[1 - \beta_i] = \underset{ALG\ j}{\operatorname{argmin}} \left[1 - \frac{P_i^{NODE} * M_{ALG\ j}^{DATA\ PER\ NODE}}{M_i^{RAM\ PER\ NODE}} \right]$$

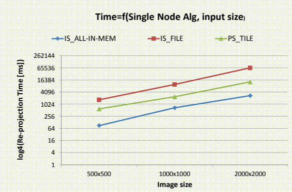
Model-Based and Measured Results

Model-Based Results





Measured Results



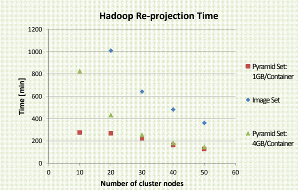


Image set to image set re-projection benchmarks for three sizes of 3D volumes and three designs of algorithms.

Two re-projection benchmarks as a function on the number of cluster nodes (Image set to image set and pyramid set to pyramid set). The benchmarks are obtained on the entire Moche 3D volume.

Summary

Our memory complexity analysis and measured time benchmarks documented that (a) the Hadoop re-projection algorithms enabled handling terabyte sized images on computer cluster/cloud platforms, and (b) the Hadoop algorithm operating on a pyramid set representation achieves similar utilization of RAM and processors per node as the two algorithms designed for a single computer operating either on pyramid set (PS_TILE FILE) or on frame set (IS_FILE).

Our contributions lie in

(a) designing a new approach to 3D re-projection of terabyte-sized images from a set of Deep Zoom pyramids,

(b) characterizing computational memory complexities of five re-projection algorithms, and

(c) maximizing the speed-up of 3D re-projection computations on Hadoop computer clusters by optimal selection of configuration parameters.

References:

[1] Deep Zoom Silverlight, Microsoft Developer Network (MSDN), URL: http://msdn.microsoft.com/en-us/library/cc645050%28VS.95%29.aspx

NIST National Institute of Standards and Technology U.S. Department of Commerce