

# Fast Parallel Image Processing on a Cluster of Workstations for Analyzing Very Large Medical Image Data Sets

Jayaram K. Udupa<sup>\*a</sup>, George J. Grevera<sup>a,b</sup>, Dewey Odhner<sup>a</sup>, Ying Zhuge<sup>a</sup>, Andre Souza<sup>a</sup>,  
Shipra Mishra<sup>a</sup>, and Tad Iwanaga<sup>a</sup>

<sup>a</sup>Medical Image Processing Group – Department of Radiology  
University of Pennsylvania – Philadelphia, PA, USA

<sup>b</sup>Department of Mathematics and Computer Science  
Saint Joseph's University – Philadelphia, PA, USA

**Purpose:** Some of the key image analysis operations such as segmentation, registration, filtering, and interpolation become choked when presented with large data sets that are now produced routinely in radiological exams. This paper presents novel strategies for parallelizing algorithms and their implementations in a portable manner so that these operations yield practical run time inexpensively even on very large data sets.

**Methods and Materials:** A cluster of workstations (*COW*) is built by interconnecting inexpensive *PCs* via a fast inexpensive 1Gbit/sec switch. Three types of parallel algorithms were devised and implemented on a *COW*. The image is divided into a set of chunks, where each chunk is a set of contiguous slices, and each chunk is assigned to each *PC* in the *COW*. In Type-1 operations, no communication is needed between *PCs* and a high level of parallelism is achieved. In Type-2 operations, inter *PC* communication is needed while a chunk is being processed. Type-3 operations are similar to Type-1, but *post hoc*, an additional set of operations are called for to complete the image analysis operation. Three image data sets – regular size (6 MB), large (241 MB), and super size (873 MB) – were utilized to test the performance of the system. These implementations were also compared with the popular open source system *ITK* in both sequential and parallel mode on a set of 10 key image analysis operations.

**Results:** In both sequential and parallel versions, the new implementations were about 2-12 times faster than *ITK* running on the same platform. Many of the *ITK* operations failed for the super data set. Even for the super data set, the most time consuming operation could be completed in under 30 minutes in the new system.

**Conclusions:** *COWs* are far more cost- and speed-effective than multiprocessing systems. The recognition of the three types of image operations helps in systematizing parallelization efforts. Portable open source systems can be built using *COWs* that are far more efficient than existing systems like *ITK*.

**Clinical Relevance/Application:** Clinical Radiological research and practice are becoming increasingly quantitative. Further, images are rapidly increasing in size and volume. For quantitative Radiology to become practical, it is crucial that the key image analysis operations such as segmentation, registration, interpolation, and filtering have algorithms and their implementations that are rapid and that yield practical run time on very large data sets. This paper describes practical and cost-effective solutions to these problems.