1 Large-Scale Data Visualization

Data being acquired and refined by a group at Johns Hopkins University (JHU) creates a new analysis and visualization challenge that many in the biomedical domain are experiencing. They have developed a complex, novel process based on 2-photon microscopy which can image organs at unprecedented levels of detail, and are using it to investigate cardiac conditions and treatments, such as cardiac ablation. Figure 1 shows a couple examples of the resolutions obtained from their imaging process, along with some post-processing we have done in an effort to draw out some of the features. However, due to the intricacy of the operation, it takes almost 3 weeks to go from a specimen to a complete, imaged heart from which they can gain insight. A full week of this process is dedicated to processing and filtering the data to reconstruct a single coherent volume from a multitude of independent samples. Worse, the data are so large that most existing visualization software is incapable of visualizing these data.

The groups standing solution was—after all the aforementioned effort to acquire these high-resolution scans—to downsample the data into a size which is amenable to current visualization software. They turned to us when they heard about our ImageVis3D tool, whose out-of-core processing capabilities decouple data size from available memory, allowing terabytes of data to be visualized on a single workstation [2]. Unfortunately our initial forays into volume visualization with these data exhibit artifacts which inhibit one's ability to pull out features of interest, as demonstrated by Figure 2.

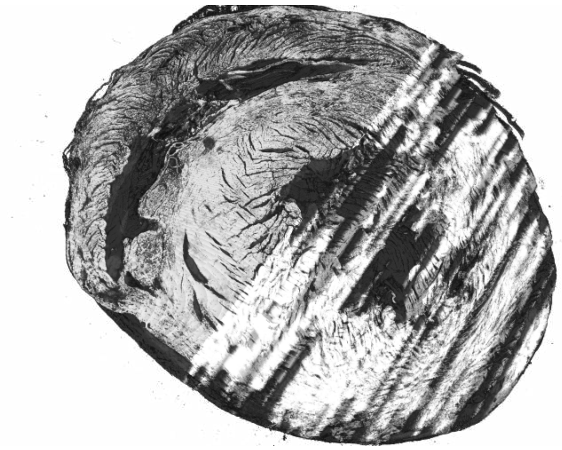

Figure 2: Heart data from the *Winslow* collaboration with a diagonal cut on one end. Collaborators wish to visualize the striations visible in the transverse plane across the coronal or sagittal planes, however the noise overcomes the signal as the cutting plane's angle of incidence increases.

The current volume reconstruction process is based on the algorithm of Gopinath et al. [4], using local window information from the scans so that the solution remains computationally tractable. Our collaborators note

> Our application to a range of 70 μm and the inherently large light scattering effect of cardiac tissue being imaged in our work created a more significant attenuation than originally considered in the design of [Gopinath et al.[4]].

To give context to this quote, portions of the heart are shaved off and the remaining heart volume is imaged. As a thick volume, energy is absorbed when it hits the tissue, but it also penetrates the volume and is absorbed and scattered at varying rates, based on the type of cells which it intersects. As such, the effect is one of a gradual as opposed to immediate attenuation, and features much larger than the desired slice thickness will accumu-

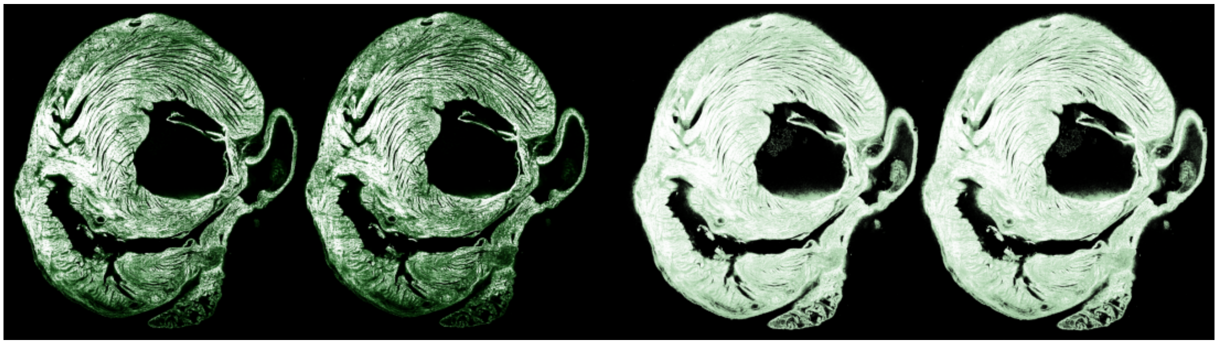

Figure 1: Preliminary image processing effects on heart data from JHU.

late in a single slice. The reconstruction process aims to minimize this effect, however as noted above, this process is leaves much to be desired.

The group's data issues are representative of a growing trend in large-scale biomedical data acquisition: large-scale acquisition requires computationally intense reconstruction processes which can take as much time as the initial scanning process. For many, the subsequent analysis processes take more time than the data acquisition itself, and thus they become the bottleneck in developing and testing biomedical hypotheses.

Increasingly, visualization and analysis software must be specifically architected to handle the scale of the data acquired.

2 Volume Visualization

We propose the research and development of biomedical and data analysis software which can help to alleviate the data-intensive science problems that groups like the one at JHU are experiencing. As leaders in the field of ray-guided volume visualization [5, 3], we are well-poised to tackle this challenge.

Ray-guided volume visualization has the potential to starkly reduce the amount of data which must be processed [1, 5, 3] to create a visualization. In 2012, Hadwiger, Pfister, et al. demonstrated that such preprocessing steps can be fruitfully applied *during* rendering, if appropriately guided by what the user needs to see. Figure 3 illustrates this concept: here, the 13 gigabyte Visible Human male has less than 500 megabytes of data visible at any one time. By explicitly considering the viewport we are outputting to, data can be reduced substantially, as shown in Figure 4.

We intend to include visualization much earlier in the processing pipeline. By using the visual analysis step to guide the reconstruction, we can produce tools which are able to provide a visual representation of the data *as* scanning proceeds, due to the data reduction implied by these ray-guided techniques.

Our proposed modified system architecture appears in Figure 5. The prominent feature is the limiting of data processing to the sections of the data which are known to be visible. We note that the 'Data store' may be any resource; classical approaches utilize local or network-attached storage here. However, we intend for this to be a network server, to further decouple visualization from data access, storage, and the reconstruction process. This allows any reconstruction process to run completely independent from the visualiza-

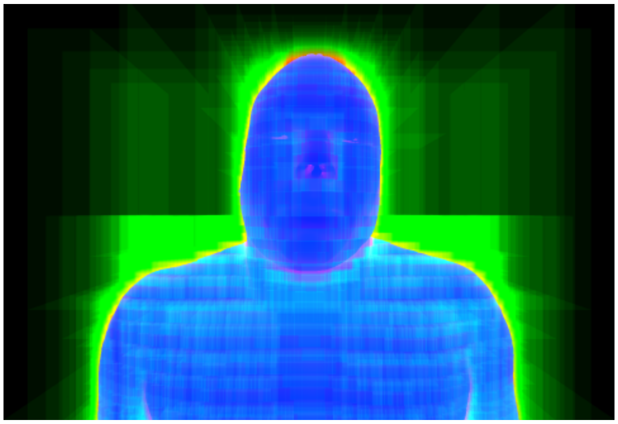

Figure 3: Per-ray behavior while rendering a large biomedical data set. Green represents space which is skipped. In red regions, dense sampling was required. Blue areas indicate that a ray terminated very quickly. Most of the rendering is blue or green, indicating that very little work needed to be done.

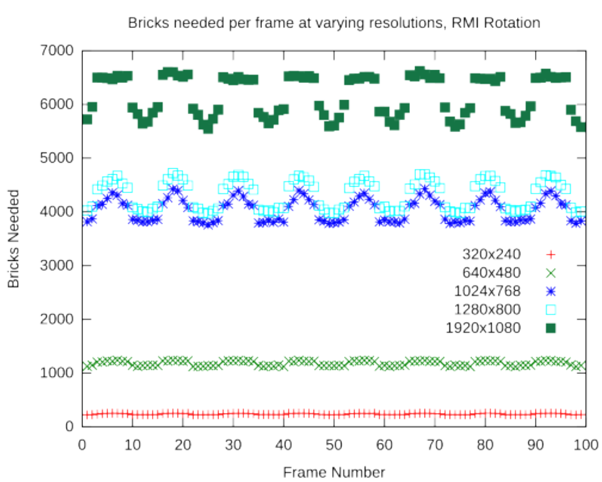

Figure 4: Amount of data needed, per-frame, to generate volume renderings of a Richtmyer-Meshkov instability at different output resolutions. Output-sensitive adaptive sampling can significantly effect the footprint of data needed for a visualization.

tion software itself, enabling truly distributed visualization. While the reconstruction process runs, the server should return any intermediate results it can—likely specific to the application in question—with an indication that results are preliminary.

References

[1] Crassin, C., Neyret, F., Lefebvre, S., and Eisemann, E. Gigavoxels: Ray-guided streaming for efficient and detailed voxel rendering. In *ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D)* (Boston, MA, Etats-Unis, feb 2009), ACM, ACM Press. `http://maverick.inria.fr/Publications/2009/CNLE09`.

[2] Fogal, T., and Krüger, J. Tuvok, an Architecture for Large Scale Volume Rendering. In *Proceedings of the 15th International Workshop on Vision, Modeling, and Visualization* (November 2010). `http://www.sci.utah.edu/~tfogal/academic/tuvok/Fogal-Tuvok.pdf`.

[3] Fogal, T., Schiewe, A., and Krüger, J. An Analysis of Scalable GPU-Based Ray-Guided Volume Rendering. `http://www.sci.utah.edu/~tfogal/academic/ray-guided/Fogal-Analysis.pdf`.

[4] Gopinath, S., Wen, Q., Thakoor, N., Luby-Phelps, K., and Gao, J. A statistical approach for intensity loss compensation of confocal microscopy images. *Journal Microscopy 230*, 143–159.

[5] Hadwiger, M., Beyer, J., Jeong, W.-K., and Pfister, H. Interactive volume exploration of petascale microscopy data streams usign a

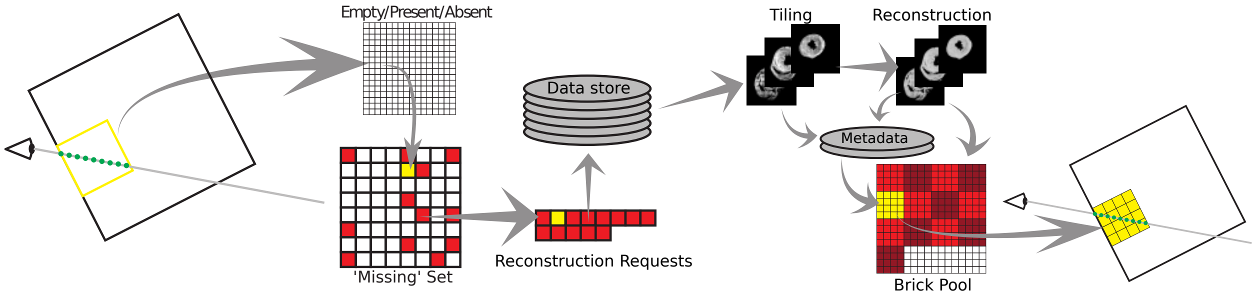

Figure 5: Visualization-guided data access pipeline. Visualization first populates a set of missing data, which is used to dictate subsets of the data to reconstruct. These subregions are queued for the reconstruction process.

visualization-driven virtual memory approach. In *Proceedings of IEEE Visualization 2012* (2012).